

THE ESSENTIAL PHYSICS OF MEDICAL IMAGING

Second Edition



**JERROLD T. BUSHBERG
J. ANTHONY SEIBERT
EDWIN M. LEIDHOLDT, JR.
JOHN M. BOONE**

THE ESSENTIAL PHYSICS OF MEDICAL IMAGING

SECOND EDITION

THE ESSENTIAL PHYSICS OF MEDICAL IMAGING

SECOND EDITION

JERROLD T. BUSHBERG, PHD

*Clinical Professor of Radiology
University of California, Davis
Sacramento, California*

J. ANTHONY SEIBERT, PHD

*Professor of Radiology
University of California, Davis
Sacramento, California*

EDWIN M. LEIDHOLDT, JR., PHD

*Associate Clinical Professor of Radiology
University of California, Davis
Sacramento, California*

JOHN M. BOONE, PHD

*Professor of Radiology
University of California, Davis
Sacramento, California*



LIPPINCOTT WILLIAMS & WILKINS

A **Wolters Kluwer** Company

Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

Acquisitions Editor: Joyce-Rachel John
Developmental Editor: Anne Snyder
Production Editor: Tony DeGeorge
Manufacturing Manager: Tim Reynolds
Cover Designer: John M. Boone and Jerrold T. Bushberg
Compositor: Lippincott Williams & Wilkins Desktop Division
Printer: Edwards Brothers

© 2002 by LIPPINCOTT WILLIAMS & WILKINS
530 Walnut Street
Philadelphia, PA 19106 USA
LWW.com

All rights reserved. This book is protected by copyright. No part of this book may be reproduced in any form or by any means, including photocopying, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotations embodied in critical articles and reviews. Materials appearing in this book prepared by individuals as part of their official duties as U.S. government employees are not covered by the above-mentioned copyright.

Printed in the USA

Library of Congress Cataloging-in-Publication Data

The essential physics of medical imaging / Jerrold T. Bushberg [et al.].-2nd ed.
p. ; cm.

Includes bibliographical references and index.

ISBN 0-683-30118-7

1. Diagnostic imaging. 2. Medical physics. I. Bushberg, Jerrold T.

[DNLM: 1. Diagnostic Imaging-methods. WN 200 E78 2001]

RC 78.7.D53 E87 2001

616.07'54-dc21

2001041711

Care has been taken to confirm the accuracy of the information presented and to describe generally accepted practices. However, the authors, editors, and publisher are not responsible for errors or omissions or for any consequences from application of the information in this book and make no warranty, expressed or implied, with respect to the currency, completeness, or accuracy of the contents of the publication. Application of this information in a particular situation remains the professional responsibility of the practitioner.

The authors, editors, and publisher have exerted every effort to ensure that drug selection and dosage set forth in this text are in accordance with current recommendations and practice at the time of publication. However, in view of ongoing research, changes in government regulations, and the constant flow of information relating to drug therapy and drug reactions, the reader is urged to check the package insert for each drug for any change in indications and dosage and for added warnings and precautions. This is particularly important when the recommended agent is a new or infrequently employed drug.

Some drugs and medical devices presented in this publication have Food and Drug Administration (FDA) clearance for limited use in restricted research settings. It is the responsibility of the health care provider to ascertain the FDA status of each drug or device planned for use in their clinical practice.

10 9 8 7 6 5 4

To my wife, Lori, who has blessed my life with her love, friendship and two marvelous children, Alexander and Jennifer, to whom this book is also dedicated. Their youthful curiosity and passion remind me that every day is filled with wonderment and discovery.

My family's love, patience, and support were essential to the completion of this book.

J.T.B.

To Spoon and T-spoon . . . for your support, patience, and understanding . . . again.

J.A.S.

To my family, especially my grandmother Mrs. Pearl Ellett Crowgey, and my teachers, especially my high school mathematics teacher Mrs. Neola Waller, and Drs. James L. Kelly, Roger Rydin, W. Reed Johnson, and Denny D. Watson of the University of Virginia.

E.M.L.

My nuclear family, mother Marion, father Gerald, and siblings Dick, Bob, and Patt, nurtured me as a babe and showed me the potential of education, the power of ambition, and the value of creativity. Susan, my wife, inspires me today as she did when I was a young man. Her affection puts the wind in my sails and the stars in my night. Daughter Emily and son Julian challenge me to be a better person every day, to view ideas from the perspective of youth, and to stop and smell the roses. Their beautiful faces and gentle souls reassure me that the future is bright. I dedicate my efforts on this book to these people, my family, my steady compatriots on this journey that is life.

J.M.B.

*There has been an
Alarming Increase
? in the Number
of Things
I Know
Nothing About*



CONTENTS

Preface xv
Acknowledgments xvii
Foreword xix

SECTION I: BASIC CONCEPTS 1

Chapter 1: Introduction to Medical Imaging 3

- 1.1 The Modalities 4
- 1.2 Image Properties 13

Chapter 2: Radiation and the Atom 17

- 2.1 Radiation 17
- 2.2 Structure of the Atom 21

Chapter 3: Interaction of Radiation with Matter 31

- 3.1 Particle Interactions 31
- 3.2 X- and Gamma Ray Interactions 37
- 3.3 Attenuation of X- and Gamma Rays 45
- 3.4 Absorption of Energy from X- and Gamma Rays 52
- 3.5 Imparted Energy, Equivalent Dose, and Effective Dose 56

Chapter 4: Computers in Medical Imaging 61

- 4.1 Storage and Transfer of Data in Computers 61
- 4.2 Analog Data and Conversion between Analog and Digital Forms 66
- 4.3 Components and Operation of Computers 70
- 4.4 Performance of Computer Systems 78
- 4.5 Computer Software 79
- 4.6 Storage, Processing, and Display of Digital Images 82

SECTION II: DIAGNOSTIC RADIOLOGY 95

Chapter 5: X-ray Production, X-ray Tubes, and Generators 97

- 5.1 Production of X-rays 97

- 5.2 X-ray Tubes 102
- 5.3 X-ray Tube Insert, Tube Housing, Filtration, and Collimation 113
- 5.4 X-ray Generator Function and Components 116
- 5.5 X-ray Generator Circuit Designs 124
- 5.6 Timing the X-ray Exposure in Radiography 132
- 5.7 Factors Affecting X-ray Emission 135
- 5.8 Power Ratings and Heat Loading 137
- 5.9 X-ray Exposure Rating Charts 140

Chapter 6: Screen-Film Radiography 145

- 6.1 Projection Radiography 145
- 6.2 Basic Geometric Principles 146
- 6.3 The Screen-Film Cassette 148
- 6.4 Characteristics of Screens 149
- 6.5 Characteristics of Film 157
- 6.6 The Screen-Film System 163
- 6.7 Contrast and Dose in Radiography 164
- 6.8 Scattered Radiation in Projection Radiography 166

Chapter 7: Film Processing 175

- 7.1 Film Exposure 175
- 7.2 The Film Processor 178
- 7.3 Processor Artifacts 181
- 7.4 Other Considerations 183
- 7.5 Laser Cameras 184
- 7.6 Dry Processing 184
- 7.7 Processor Quality Assurance 186

Chapter 8: Mammography 191

- 8.1 X-ray Tube Design 194
- 8.2 X-ray Generator and Phototimer System 204
- 8.3 Compression, Scattered Radiation, and Magnification 207
- 8.4 Screen-Film Cassettes and Film Processing 212
- 8.5 Ancillary Procedures 219
- 8.6 Radiation Dosimetry 222
- 8.7 Regulatory Requirements 224

Chapter 9: Fluoroscopy 231

- 9.1 Functionality 231
- 9.2 Fluoroscopic Imaging Chain Components 232
- 9.3 Peripheral Equipment 242
- 9.4 Fluoroscopy Modes of Operation 244
- 9.5 Automatic Brightness Control (ABC) 246
- 9.6 Image Quality 248
- 9.7 Fluoroscopy Suites 249
- 9.8 Radiation Dose 251

Chapter 10: Image Quality 255

- 10.1 Contrast 255
- 10.2 Spatial Resolution 263
- 10.3 Noise 273
- 10.4 Detective Quantum Efficiency (DQE) 283
- 10.5 Sampling and Aliasing in Digital Images 283
- 10.6 Contrast-Detail Curves 287
- 10.7 Receiver Operating Characteristics Curves 288

Chapter 11: Digital Radiography 293

- 11.1 Computed Radiography 293
- 11.2 Charged-Coupled Devices (CCDs) 297
- 11.3 Flat Panel Detectors 300
- 11.4 Digital Mammography 304
- 11.5 Digital versus Analog Processes 307
- 11.6 Implementation 307
- 11.7 Patient Dose Considerations 308
- 11.8 Hard Copy versus Soft Copy Display 308
- 11.9 Digital Image Processing 309
- 11.10 Contrast versus Spatial Resolution in Digital Imaging 315

Chapter 12: Adjuncts to Radiology 317

- 12.1 Geometric Tomography 317
- 12.2 Digital Tomosynthesis 320
- 12.3 Temporal Subtraction 321
- 12.4 Dual-Energy Subtraction 323

Chapter 13: Computed Tomography 327

- 13.1 Basic Principles 327
- 13.2 Geometry and Historical Development 331
- 13.3 Detectors and Detector Arrays 339
- 13.4 Details of Acquisition 342
- 13.5 Tomographic Reconstruction 346
- 13.6 Digital Image Display 358
- 13.7 Radiation Dose 362
- 13.8 Image Quality 367
- 13.9 Artifacts 369

Chapter 14: Nuclear Magnetic Resonance 373

- 14.1 Magnetization Properties 373
- 14.2 Generation and Detection of the Magnetic Resonance Signal 381
- 14.3 Pulse Sequences 391
- 14.4 Spin Echo 391
- 14.5 Inversion Recovery 399
- 14.6 Gradient Recalled Echo 403
- 14.7 Signal from Flow 408
- 14.8 Perfusion and Diffusion Contrast 409
- 14.9 Magnetization Transfer Contrast 411

Chapter 15: Magnetic Resonance Imaging (MRI) 415

- 15.1 Localization of the MR Signal 415
- 15.2 k-space Data Acquisition and Image Reconstruction 426
- 15.3 Three-Dimensional Fourier Transform Image Acquisition 438
- 15.4 Image Characteristics 439
- 15.5 Angiography and Magnetization Transfer Contrast 442
- 15.6 Artifacts 447
- 15.7 Instrumentation 458
- 15.8 Safety and Bioeffects 465

Chapter 16: Ultrasound 469

- 16.1 Characteristics of Sound 470
- 16.2 Interactions of Ultrasound with Matter 476

16.3	Transducers	483
16.4	Beam Properties	490
16.5	Image Data Acquisition	501
16.6	Two-Dimensional Image Display and Storage	510
16.7	Miscellaneous Issues	516
16.8	Image Quality and Artifacts	524
16.9	Doppler Ultrasound	531
16.10	System Performance and Quality Assurance	544
16.11	Acoustic Power and Bioeffects	548
Chapter 17:	Computer Networks, PACS, and Teleradiology	555
17.1	Computer Networks	555
17.2	PACS and Teleradiology	565
SECTION III: NUCLEAR MEDICINE 587		
Chapter 18:	Radioactivity and Nuclear Transformation	589
18.1	Radionuclide Decay Terms and Relationships	589
18.2	Nuclear Transformation	593
Chapter 19:	Radionuclide Production and Radiopharmaceuticals	603
19.1	Radionuclide Production	603
19.2	Radiopharmaceuticals	617
19.3	Regulatory Issues	624
Chapter 20:	Radiation Detection and Measurement	627
20.1	Types of Detectors	627
20.2	Gas-Filled Detectors	632
20.3	Scintillation Detectors	636
20.4	Semiconductor Detectors	641
20.5	Pulse Height Spectroscopy	644
20.6	Non-Imaging Detector Applications	654
20.7	Counting Statistics	661
Chapter 21:	Nuclear Imaging—The Scintillation Camera	669
21.1	Planar Nuclear Imaging: The Anger Scintillation Camera	670
21.2	Computers in Nuclear Imaging	695

Chapter 22: Nuclear Imaging—Emission Tomography 703

22.1 Single Photon Emission Computed Tomography (SPECT) 704

22.2 Positron Emission Tomography (PET) 719

SECTION IV: RADIATION PROTECTION, DOSIMETRY, AND BIOLOGY 737

Chapter 23: Radiation Protection 739

23.1 Sources of Exposure to Ionizing Radiation 739

23.2 Personnel Dosimetry 747

23.3 Radiation Detection Equipment in Radiation Safety 753

23.4 Radiation Protection and Exposure Control 755

23.5 Regulatory Agencies and Radiation Exposure Limits 788

Chapter 24: Radiation Dosimetry of the Patient 795

24.1 X-ray Dosimetry 800

24.2 Radiopharmaceutical Dosimetry: The MIRD Method 805

Chapter 25: Radiation Biology 813

25.1 Interaction of Radiation with Tissue 814

25.2 Cellular Radiobiology 818

25.3 Response of Organ Systems to Radiation 827

25.4 Acute Radiation Syndrome 831

25.5 Radiation-Induced Carcinogenesis 838

25.6 Hereditary Effects of Radiation Exposure 851

25.7 Radiation Effects *In Utero* 853

SECTION V: APPENDICES 863

Appendix A: Fundamental Principles of Physics 865

A.1 Physical Laws, Quantities, and Units 865

A.2 Classical Physics 867

A.3 Electricity and Magnetism 868

Appendix B: Physical Constants, Prefixes, Geometry, Conversion Factors, and Radiologic Data 883

B.1 Physical Constants, Prefixes, and Geometry 883

B.2 Conversion Factors 884

B.3 Radiological Data for Elements 1 through 100 885**Appendix C: Mass Attenuation Coefficients and Spectra Data Tables 887**

- C.1 Mass Attenuation Coefficients for Selected Elements 887
- C.2 Mass Attenuation Coefficients for Selected Compounds 889
- C.3 Mass Energy Attenuation Coefficients for Selected Detector Compounds 890
- C.4 Mammography Spectra: Mo/Mo 891
- C.5 Mammography Spectra: Mo/Rh 893
- C.6 Mammography Spectra: Rh/Rh 895
- C.7 General Diagnostic Spectra: W/Al 897

Appendix D: Radiopharmaceutical Characteristics and Dosimetry 899

- D.1 Route of administration, localization, clinical utility, and other characteristics of commonly used radiopharmaceuticals 900
- D.2 Typical administered adult activity, highest organ dose, gonadal dose, and adult effective dose for commonly used radiopharmaceuticals 908
- D.3 Effective doses per unit activity administered to patients age 15, 10, 5, and 1 year for commonly used diagnostic radiopharmaceuticals 910
- D.4 Absorbed dose estimates to the embryo/fetus per unit activity administered to the mother for commonly used radiopharmaceuticals 911

Appendix E: Internet Resources 913**Subject Index 915**

PREFACE TO THE SECOND EDITION

The first edition of this text was developed from the extensive syllabus we had created for a radiology resident board review course that has been taught annually at the University of California Davis since 1984. Although the topics were, in broad terms, the same as in the course syllabus, the book itself was written *de novo*. Since the first edition of this book was completed in 1993, there have been many important advances in medical imaging technology. Consequently, in this second edition, most of the chapters have been completely rewritten, although the organization of the text into four main sections remains unchanged. In addition, new chapters have been added. *An Introduction to Medical Imaging* begins this new edition as Chapter 1. In the *Diagnostic Radiology* section, chapters on *Film Processing*, *Digital Radiography*, and *Computer Networks, PACS, and Teleradiography* have been added. In recognition of the increased sophistication and complexity in some modalities, the chapters on MRI and nuclear imaging have been split into two chapters each, in an attempt to break the material into smaller and more digestible parts. Considerable effort was also spent on integrating the discussion and assuring consistent terminology between the different chapters. The *Image Quality* chapter was expanded to provide additional details on this important topic.

In addition, a more extensive set of reference data is provided in this edition. The appendices have been expanded to include the fundamental principles of physics, physical constants and conversion factors, elemental data, mass attenuation coefficients, x-ray spectra, and radiopharmaceutical characteristics and dosimetry. Web sites of professional societies, governmental organizations and other entities that may be of interest to the medical imaging community are also provided.

The field of radiology is in a protracted state of transition regarding the usage of units. Although the SI unit system has been officially adopted by most radiology and scientific journals, it is hard to avoid the use of the roentgen and rem. Our ionization chambers still read out in milliroentgen of exposure (not milligray of air kerma), and our monthly film badge reports are still conveyed in millirem (not millisieverts). The U.S. Government has been slow to adopt SI units. Consequently, while we have adopted SI units throughout most of the text, we felt compelled to discuss (and use where appropriate) the older units in contexts where they are still used. Furthermore, antiquated quantities such as the *effective dose equivalent* are still used by the U.S. Nuclear Regulatory Commission, although the rest of the world uses *effective dose*.

We have received many comments over the years from instructors, residents, and other students who made use of the first edition, and we have tried to respond to these comments by making appropriate changes in the book. Our intention with this book is to take the novice reader from the introduction of a topic, all the way through a relatively thorough description of it. If we try to do this using too few words we may lose many readers; if we use too many words we may bore others. We did our best to walk this fine line, but if you are in the latter group, we encourage you to *read faster*.

We are deeply grateful to that part of the radiology community who embraced our first effort. This second edition was inspired both by the successes and the shortcomings of the first edition. We are also grateful to those who provided suggestions for improvement and we hope that they will be pleased with this new edition.

Jerrold T. Bushberg
J. Anthony Seibert
Edwin M. Leidholdt, Jr.
John M. Boone

ACKNOWLEDGMENTS

During the production of this work, several individuals generously gave their time and expertise. First, we would like to thank L. Stephen Graham, Ph.D., University of California, Los Angeles, and Mark W. Groch, Ph.D., Northwestern University, who provided valuable insight in detailed reviews of the chapters on nuclear medicine imaging. We also thank Michael Buonocore, M.D., Ph.D., University of California, Davis, who reviewed the chapters on MRI, and Fred Mettler, M.D., University of New Mexico, who provided valuable contributions to the chapter on radiation biology. Raymond Tanner, Ph.D., University of Tennessee, Memphis, provided a useful critique and recommended changes in several chapters of the First Edition, which were incorporated into this effort. Virgil Cooper, Ph.D., University of California, Los Angeles, provided thoughtful commentary on x-ray imaging and a fresh young perspective for gauging our efforts.

We are also appreciative of the comments of Stewart Bushong, Ph.D., Baylor College of Medicine, especially regarding film processing. Walter Huda, Ph.D., SUNY Upstate Medical University, provided very helpful discussions on many topics. The expertise of Mel Tecotzky, Ph.D., in x-ray phosphors enhanced our discussion of this topic. Skip Kennedy, M.S., University of California, Davis, provided technical insight regarding computer networks and PACS.

The efforts of Fernando Herrera, UCD Illustration Services, brought to life some of the illustrations used in several chapters. In addition, we would like to acknowledge the superb administrative support of Lorraine Smith and Patrice Wilbur, whose patience and attention to detail are greatly appreciated.

We are grateful for the contributions that these individuals have made towards the development of this book. We are also indebted to many other scientists whose work in this field predates our own and whose contributions served as the foundation of many of the concepts developed in this book.

J.T.B.
J.A.S.
E.M.L.
J.M.B.

FOREWORD

Can medical physics be interesting and exciting? Personally, I find most physics textbooks dry, confusing, and a useful cure for my insomnia. This book is different. Dr. Bushberg and his colleagues have been teaching residents as well as an international review course in radiation physics, protection, dosimetry, and biology for almost two decades. They know what works, what does not, and how to present information clearly.

A particularly strong point of this book is that it covers all areas of diagnostic imaging. A number of current texts cover only one area of physics and the residents often purchase several texts by different authors in order to have a complete grasp of the subject matter.

Of course, medical imagers are more at home with pictures rather than text and formulas. Most authors of other physics books have not grasped this concept. The nearly 600 exquisite illustrations contained in this substantially revised second edition will make this book a favorite of the medical imaging community.

*Fred A. Mettler Jr., M.D.
Professor and Chair
Department of Radiology
University of New Mexico
Albuquerque, New Mexico*

SECTION
I

BASIC CONCEPTS

INTRODUCTION TO MEDICAL IMAGING

Medical imaging of the human body requires some form of energy. In the medical imaging techniques used in radiology, the energy used to produce the image must be capable of penetrating tissues. Visible light, which has limited ability to penetrate tissues at depth, is used mostly outside of the radiology department for medical imaging. Visible light images are used in dermatology (skin photography), gastroenterology and obstetrics (endoscopy), and pathology (light microscopy). Of course, all disciplines in medicine use direct visual observation, which also utilizes visible light. In diagnostic radiology, the electromagnetic spectrum outside the visible light region is used for x-ray imaging, including mammography and computed tomography, magnetic resonance imaging, and nuclear medicine. Mechanical energy, in the form of high-frequency sound waves, is used in ultrasound imaging.

With the exception of nuclear medicine, all medical imaging requires that the energy used to penetrate the body's tissues also interact with those tissues. If energy were to pass through the body and not experience some type of interaction (e.g., absorption, attenuation, scattering), then the detected energy would not contain any useful information regarding the internal anatomy, and thus it would not be possible to construct an image of the anatomy using that information. In nuclear medicine imaging, radioactive agents are injected or ingested, and it is the metabolic or physiologic *interactions* of the agent that give rise to the information in the images.

While medical images can have an aesthetic appearance, the diagnostic utility of a medical image relates to both the technical quality of the image and the conditions of its acquisition. Consequently, the assessment of image quality in medical imaging involves very little artistic appraisal and a great deal of technical evaluation. In most cases, the image quality that is obtained from medical imaging devices involves compromise—better x-ray images can be made when the radiation dose to the patient is high, better magnetic resonance images can be made when the image acquisition time is long, and better ultrasound images result when the ultrasound power levels are large. Of course, patient safety and comfort must be considered when acquiring medical images; thus excessive patient dose in the pursuit of a perfect image is not acceptable. Rather, the power levels used to make medical images require a balance between patient safety and image quality.

1.1 THE MODALITIES

Different types of medical images can be made by varying the types of energies used and the acquisition technology. The different modes of making images are referred to as *modalities*. Each modality has its own applications in medicine.

Radiography

Radiography was the first medical imaging technology, made possible when the physicist Wilhelm Roentgen discovered x-rays on November 8, 1895. Roentgen also made the first radiographic images of human anatomy (Fig. 1-1). Radiography (also called roentgenography) defined the field of radiology, and gave rise to radiologists, physicians who specialize in the interpretation of medical images. Radiography is performed with an x-ray source on one side of the patient, and a (typically flat) x-ray detector on the other side. A short duration (typically less than $\frac{1}{2}$ second) pulse of x-rays is emitted by the x-ray tube, a large fraction of the x-rays interacts in the patient, and some of the x-rays pass through the patient and reach the detector, where a radiographic image is formed. The homogeneous distribution of x-rays that enter the patient is modified by the degree to which the x-rays are removed from the beam (i.e., attenuated) by scattering and absorption within the tissues. The attenuation properties of tissues such as bone, soft tissue, and air inside the patient are very different, resulting in the heterogeneous distribution of x-rays that emerges from the patient. The radiographic image is a picture of this x-ray distribution. The detector used in radiography can be photographic film (e.g., screen-film radiography) or an electronic detector system (i.e., digital radiography).



FIGURE 1-1. The beginning of diagnostic radiology, represented by this famous radiographic image made on December 22, 1895 of the wife of the discoverer of x-rays, Wilhelm Conrad Roentgen. The bones of her hand as well as two rings on her finger are clearly visible. Within a few months, Roentgen was able to determine the basic physical properties of x-rays. Nearly simultaneously, as word of the discovery spread around the world, medical applications of this "new kind of ray" propelled radiologic imaging into an essential component of medical care. In keeping with mathematical conventions, Roentgen assigned the letter "x" to represent the unknown nature of the ray and thus the term *x-ray* was born. Details regarding x-ray production and interactions can be found in Chapters 5 and 3, respectively. (Reproduced from Glasser O. *Wilhelm Conrad and Röntgen and the early history of the roentgen rays*. Springfield, IL: Charles C. Thomas, 1933, with permission.)

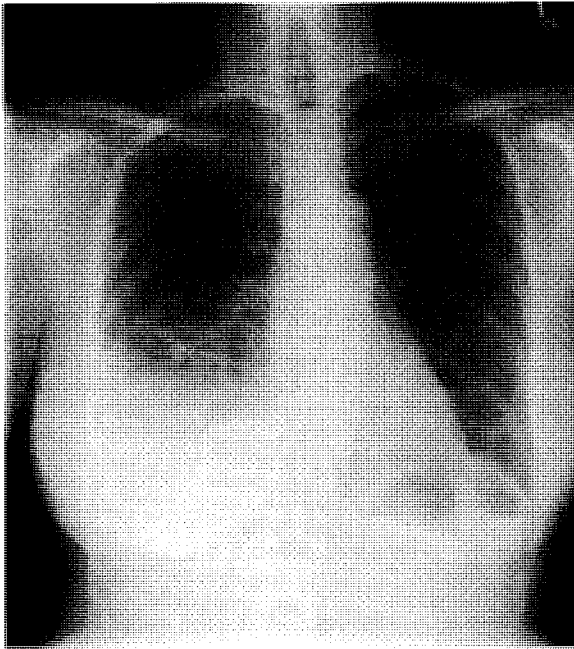


FIGURE 1-2. The chest x-ray is the most ubiquitous image in diagnostic radiology. High x-ray energy is used for the purpose of penetrating the mediastinum, cardiac, and diaphragm areas of the patient without overexposing areas within the lungs. Variation in the gray-scale image represents an attenuation map of the x-rays: dark areas (high film optical density) have low attenuation, and bright areas (low film optical density) have high attenuation. The image here shows greater than normal attenuation in the lower lobes of the lungs, consistent with plural effusion, right greater than left. Rapid acquisition, low risk, low cost, and high diagnostic value are the major reasons why x-ray projection imaging represents the bulk of all diagnostic imaging studies (typically 60% to 70% of all images produced). Projection imaging physics is covered in Chapter 6.

Transmission imaging refers to imaging in which the energy source is outside the body on one side, and the energy passes through the body and is detected on the other side of the body. Radiography is a transmission imaging modality. *Projection imaging* refers to the case when each point on the image corresponds to information along a straight line trajectory through the patient. Radiography is also a projection imaging modality. Radiographic images are useful for a very wide range of medical indications, including the diagnosis of broken bones, lung cancer, cardiovascular disorders, etc. (Fig. 1-2).

Fluoroscopy

Fluoroscopy refers to the continuous acquisition of a sequence of x-ray images over time, essentially a real-time x-ray movie of the patient. Fluoroscopy is a transmission projection imaging modality, and is, in essence, just real-time radiography. Fluoroscopic systems use x-ray detector systems capable of producing images in rapid temporal sequence. Fluoroscopy is used for positioning catheters in arteries, for visualizing contrast agents in the gastrointestinal (GI) tract, and for other medical applications such as invasive therapeutic procedures where real-time image feedback is necessary. Fluoroscopy is also used to make x-ray movies of anatomic motion, such as of the heart or the esophagus.

Mammography

Mammography is radiography of the breast, and is thus a transmission projection type of imaging. Much lower x-ray energies are used in mammography than any other radiographic applications, and consequently modern mammography uses



FIGURE 1-3. Mammography is a specialized x-ray projection imaging technique useful for detecting breast anomalies such as masses and calcifications. The mammogram in the image above demonstrates a spiculated mass (*arrow*) with an appearance that is typical of a cancerous lesion in the breast, in addition to blood vessels and normal anatomy. Dedicated mammography equipment using low x-ray energies, k-edge filters, compression, screen/film detectors, antiscatter grids, and automatic exposure control results in optimized breast images of high quality and low x-ray dose, as detailed in Chapter 8. X-ray mammography is the current procedure of choice for screening and early detection of breast cancer because of high sensitivity, excellent benefit to risk, and low cost.

x-ray machines and detector systems specifically designed for breast imaging. Mammography is used to screen asymptomatic women for breast cancer (screening mammography), and is also used to help in the diagnosis of women with breast symptoms such as the presence of a lump (diagnostic mammography) (Fig. 1-3).

Computed Tomography (CT)

CT became clinically available in the early 1970s and is the first medical imaging modality made possible by the computer. CT images are produced by passing x-rays through the body, at a large number of angles, by rotating the x-ray tube around the body. One or more linear detector arrays, opposite the x-ray source, collect the transmission projection data. The numerous data points collected in this manner are synthesized by a computer into a *tomographic* image of the patient. The term *tomography* refers to a picture (*-graph*) of a slice (*tomo-*). CT is a transmission technique that results in images of individual slabs of tissue in the patient. The advantage of a tomographic image over projection image is its ability to display the anatomy in a slab (slice) of tissue in the absence of over- or underlying structures.

CT changed the practice of medicine by substantially reducing the need for exploratory surgery. Modern CT scanners can acquire 5-mm-thick tomographic images along a 30-cm length of the patient (i.e., 60 images) in 10 seconds, and reveal the presence of cancer, ruptured discs, subdural hematomas, aneurysms, and a large number of other pathologies (Fig. 1-4).



FIGURE 1-4. A computed tomography (CT) image of the abdomen reveals a ruptured disc (arrow) manifested as the bright area of the image adjacent to the vertebral column. Anatomic structures such as the kidneys, arteries, and intestines are clearly represented in the image. CT provides high-contrast sensitivity for soft tissue, bone, and air interfaces without superimposition of anatomy. With recently implemented multiple array detectors, scan times of 0.5 seconds per 360 degrees and fast computer reconstruction permits head-to-toe imaging in as little as 30 seconds. Because of fast acquisition speed, high-contrast sensitivity, and ability to image tissue, bone, and air, CT remains the workhorse of tomographic imaging in diagnostic radiology. Chapter 13 describes the details of CT.

Nuclear Medicine Imaging

Nuclear medicine is the branch of radiology in which a chemical or compound containing a radioactive isotope is given to the patient orally, by injection, or by inhalation. Once the compound has distributed itself according to the physiologic status of the patient, a radiation detector is used to make projection images from the x- and/or gamma rays emitted during radioactive decay of the agent. Nuclear medicine produces *emission* images (as opposed to transmission images), because the radioisotopes emit their energy from inside the patient.

Nuclear medicine imaging is a form of *functional* imaging. Rather than yielding information about just the anatomy of the patient, nuclear medicine images provide information regarding the physiologic conditions in the patient. For example, thallium tends to concentrate in normal heart muscle, but does not concentrate as well in areas that are infarcted or ischemic. These areas appear as “cold spots” on a nuclear medicine image, and are indicative of the functional status of the heart. Thyroid tissue has a great affinity for iodine, and by administering radioactive iodine (or its analogs), the thyroid can be imaged. If thyroid cancer has metastasized in the patient, then “hot spots” indicating their location will be present on the nuclear medicine images. Thus functional imaging is the forte of nuclear medicine.

Nuclear Medicine Planar Imaging

Nuclear medicine planar images are projection images, since each point on the image is representative of the radioisotope activity along a line projected through the patient. Planar nuclear images are essentially two-dimensional maps of the radioisotope distribution, and are helpful in the evaluation of a large number of disorders (Fig. 1-5).

Single Photon Emission Computed Tomography (SPECT)

SPECT is the tomographic counterpart of nuclear medicine planar imaging, just like CT is the tomographic counterpart of radiography. In SPECT, a nuclear camera records x- or gamma-ray emissions from the patient from a series of different

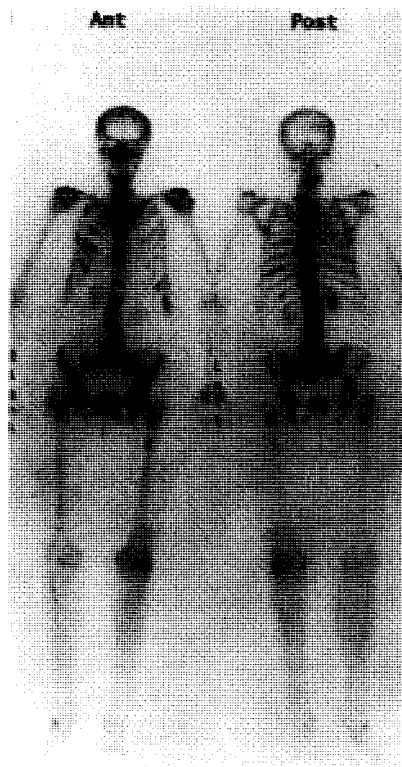


FIGURE 1-5. Anterior and posterior whole-body bone scan of a 74-year-old woman with a history of right breast cancer. This patient was injected with 925 MBq (25 mCi) of technetium (Tc) 99m methylenediphosphonate (MDP) and was imaged 3 hours later with a dual-headed whole-body scintillation camera. The scan demonstrates multiple areas of osteoblastic metastases in the axial and proximal skeleton. Incidental findings include an arthritis pattern in the shoulders and left knee. Computer processed planar imaging is still the standard for many nuclear medicine examinations (e.g., whole-body bone scans, hepatobiliary, thyroid, renal, and pulmonary studies). Planar nuclear imaging is discussed in detail in Chapter 21. (Image courtesy of Dr. David Shelton, University of California, Davis Medical Center, Davis, CA.)

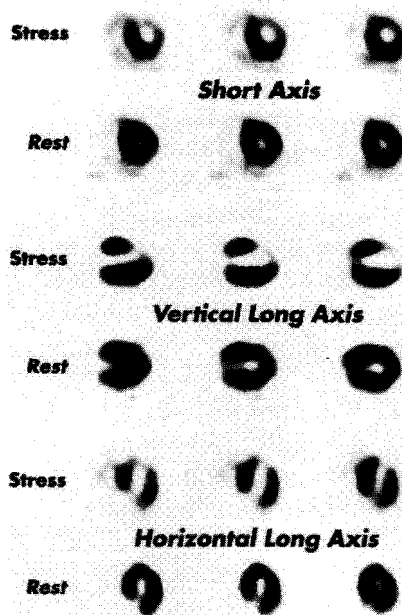


FIGURE 1-6. A myocardial perfusion stress test utilizing thallium 201 (Tl 201) and single photon emission computed tomography (SPECT) imaging was performed on a 79-year-old woman with chest pain. This patient had pharmacologic stress with dipyridamole and was injected with 111 MBq (3 mCi) of Tl 201 at peak stress. Stress imaging followed immediately on a variable-angle two-headed SPECT camera. Image data was acquired over 180 degrees at 30 seconds per stop. The rest/redistribution was done 3 hours later with a 37-MBq (1-mCi) booster injection of Tl 201. Short axis, horizontal long axis, and vertical long axis views show relatively reduced perfusion on anterolateral wall stress images, with complete reversibility on rest/redistribution images. Findings indicated coronary stenosis in the left anterior descending (LAD) coronary artery distribution. SPECT is now the standard for a number of nuclear medicine examinations including cardiac perfusion and brain and tumor imaging. SPECT imaging is discussed in detail in Chapter 22. (Image courtesy of Dr. David Shelton, University of California, Davis Medical Center, Davis, CA.)

angles around the patient. These projection data are used to reconstruct a series of tomographic emission images. SPECT images provide diagnostic functional information similar to nuclear planar examinations; however, their tomographic nature allows physicians to better understand the precise distribution of the radioactive agent, and to make a better assessment of the function of specific organs or tissues within the body (Fig. 1-6). The same radioactive isotopes are used in both planar nuclear imaging and SPECT.

Positron Emission Tomography (PET)

Positrons are positively charged electrons, and are emitted by some radioactive isotopes such as fluorine 18 and oxygen 15. These radioisotopes are incorporated into metabolically relevant compounds [such as ^{18}F -fluorodeoxyglucose (FDG)], which localize in the body after administration. The decay of the isotope produces a positron, which rapidly undergoes a very unique interaction: the positron (e^+) combines with an electron (e^-) from the surrounding tissue, and the mass of both the e^+ and the e^- is converted by annihilation into pure energy, following Einstein's famous equation $E = mc^2$. The energy that is emitted is called *annihilation radiation*. Annihilation radiation production is similar to gamma-ray emission, except that *two* photons are emitted, and they are emitted in almost exactly opposite directions, i.e., 180 degrees from each other. A PET scanner utilizes a ring of detectors that surround the patient, and has special circuitry that is capable of identifying the photon pairs produced during annihilation. When a photon pair is detected by two detectors on the scanner, it is known that the decay event took place somewhere along a straight line between those two detectors. This information is used to mathematically compute the three-dimensional distribution of the PET agent, resulting in a series of tomographic emission images.

Although more expensive than SPECT, PET has clinical advantages in certain diagnostic areas. The PET detector system is more sensitive to the presence of radioisotopes than SPECT cameras, and thus can detect very subtle pathologies. Furthermore, many of the elements that emit positrons (carbon, oxygen, fluorine) are quite physiologically relevant (fluorine is a good substitute for a hydroxyl group), and can be incorporated into a large number of biochemicals. The most important of these is ^{18}F FDG, which is concentrated in tissues of high glucose metabolism such as primary tumors and their metastases (Fig. 1-7).

Magnetic Resonance Imaging (MRI)

MRI scanners use magnetic fields that are about 10,000 to 60,000 times stronger than the earth's magnetic field. Most MRI utilizes the nuclear magnetic resonance properties of the proton—i.e., the nucleus of the hydrogen atom, which is very abundant in biologic tissues (each cubic millimeter of tissue contains about 10^{18} protons). The proton has a magnetic moment, and when placed in a 1.5-tesla (T) magnetic field, the proton will preferentially absorb radio wave energy at the resonance frequency of 63 megahertz (MHz).

In MRI, the patient is placed in the magnetic field, and a pulse of radio waves is generated by antennas ("coils") positioned around the patient. The protons in the patient absorb the radio waves, and subsequently reemit this radio wave energy after a period of time that depends on the very localized magnetic properties of the sur-

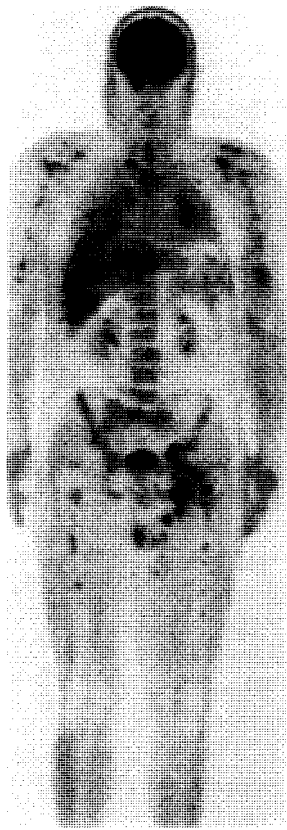


FIGURE 1-7. Whole-body positron emission tomography (PET) scan of a 54-year-old man with malignant melanoma. Patient was injected intravenously with 600 MBq (16 mCi) of ^{18}F -deoxyglucose. The patient was imaged for 45 minutes, beginning 75 minutes after injection of the radiopharmaceutical. The image demonstrates extensive metastatic disease with abnormalities throughout the axial and proximal appendicular skeleton, right and left lungs, liver, and left inguinal and femoral lymph nodes. The unique ability of the PET scan in this case was to correctly assess the extent of disease, which was underestimated by CT, and to serve as a baseline against which future comparisons could be made to assess the effects of immunochemotherapy. PET has applications in functional brain and cardiac imaging and is rapidly becoming a routine diagnostic tool in the staging of many cancers. PET technology is discussed in detail in Chapter 22. (Image courtesy of Dr. George Segall, VA Medical Center, Palo Alto, CA.)

rounding tissue. The radio waves emitted by the protons in the patient are detected by the antennas that surround the patient. By slightly changing the strength of the magnetic field as a function of position in the patient (using magnetic field *gradients*), the proton resonance frequency will vary as a function of position, since frequency is proportional to magnetic field strength. The MRI system uses the frequency (and phase) of the returning radio waves to determine the position of each signal from the patient. The mode of operation of MRI systems is often referred to as *spin echo* imaging.

MRI produces a set of tomographic slices through the patient, where each point in the image depends on the micromagnetic properties of the corresponding tissue at that point. Because different types of tissue such as fat, white, and gray matter in the brain, cerebral spinal fluid, and cancer all have different local magnetic properties, images made using MRI demonstrate high sensitivity to anatomic variations and therefore are high in contrast. MRI has demonstrated exceptional utility in neurologic imaging (head and spine), and for musculoskeletal applications such as imaging the knee after athletic injury (Fig. 1-8).

MRI is a tomographic imaging modality, and competes with x-ray CT in many clinical applications. The acquisition of the highest-quality images using MRI requires tens of minutes, whereas a CT scan of the entire head requires about 10 seconds. Thus, for patients where motion cannot be controlled (pediatric patients) or in anatomic areas where involuntary patient motion occurs (the beating heart

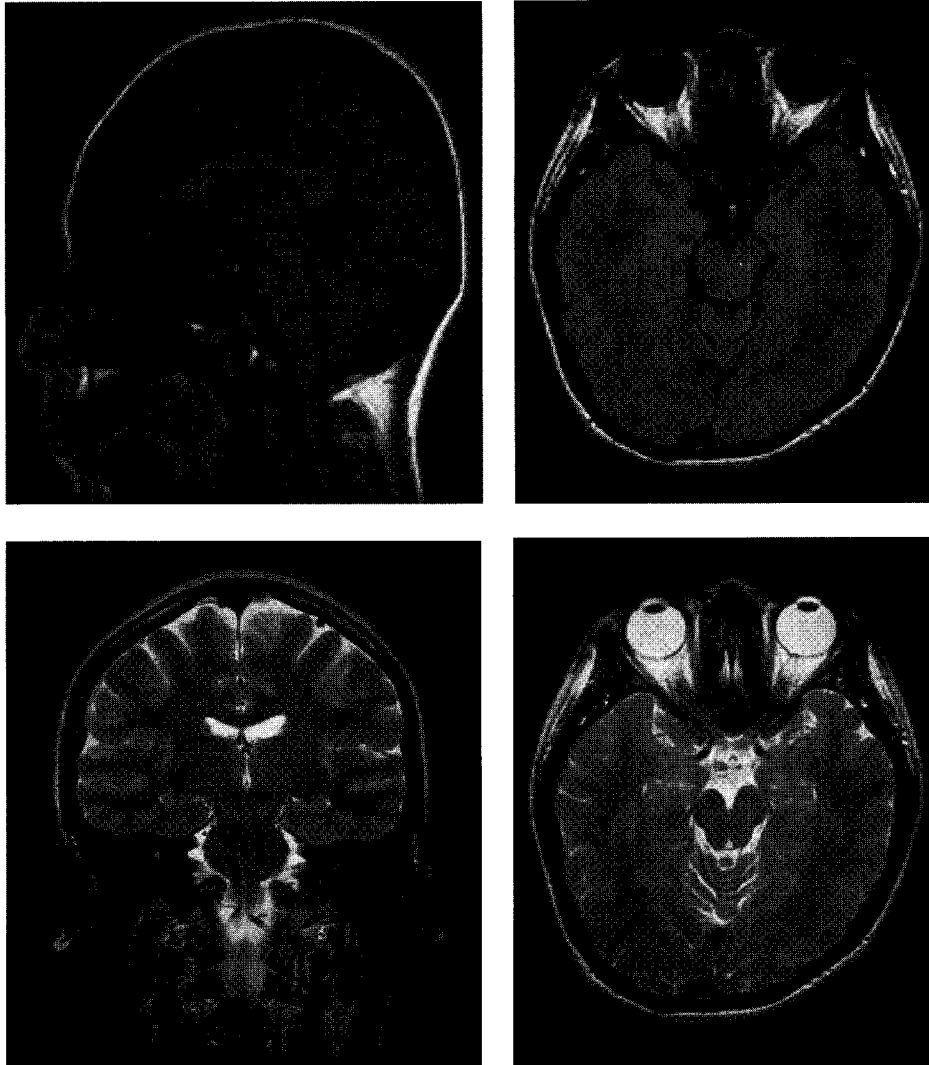


FIGURE 1-8. Sagittal (*upper left*), coronal (*lower left*), and axial (*right*) normal images of the brain demonstrate the exquisite contrast sensitivity and acquisition capability of magnetic resonance imaging (MRI). Image contrast is generated by specific MR pulse sequences (*upper images* are T1-weighted, *lower images* are T2 weighted) to emphasize the magnetization characteristics of the tissues placed in a strong magnetic field, based on selective excitation and reemission of radiofrequency signals. MRI is widely used for anatomic as well as physiologic and functional imaging, and is the modality of choice for neurologic assessment, staging of cancer, and physiologic/anatomic evaluation of extremities and joints. Further information regarding MRI can be found in Chapters 14 and 15.

and the churning intestines), CT is often used instead of MRI. Also, because of the large magnetic field used in MRI, electronic monitoring equipment cannot be used while the patient is being scanned. Thus, for most trauma, CT is preferred. MRI should not be performed on patients who have cardiac pacemakers or internal ferromagnetic objects such as surgical aneurysm clips, metal plate or rod implants, or metal shards near critical structures like the eye.

Despite some indications where MRI should not be used, fast image acquisition techniques using special coils have made it possible to produce images in much shorter periods of time, and this has opened up the potential of using MRI for imaging of the motion-prone thorax and abdomen. MRI scanners can also detect the presence of motion, which is useful for monitoring blood flow through arteries (MR *angiography*).

Ultrasound Imaging

When a book is dropped on a table, the impact causes pressure waves (called sound) to propagate through the air such that they can be heard at a distance. Mechanical energy in the form of high-frequency (“ultra”) sound can be used to generate images of the anatomy of a patient. A short-duration pulse of sound is generated by an ultrasound *transducer* that is in direct physical contact with the tissues being imaged. The sound waves travel into the tissue, and are reflected by internal structures in the body, creating echoes. The reflected sound waves then reach the trans-



FIGURE 1-9. The ultrasound image is a map of the echoes from tissue boundaries of high-frequency sound wave pulses (typically from 2 to 20 MHz frequency) gray-scale encoded into a two-dimensional tomographic image. A phased-array transducer operating at 3.5 MHz produced this normal obstetrical ultrasound image of Jennifer Lauren Bushberg (at approximately 3½ months). Variations in the image brightness are due to acoustic characteristics of the tissues; for example, the fluid in the placenta is echo free, while most fetal tissues are echogenic and produce a larger signal strength. Shadowing is caused by highly attenuating or scattering tissues such as bone or air and the corresponding distal low-intensity streaks. Besides tomographic acoustic imaging, distance measurements (e.g., fetal head diameter assessment for aging), and vascular assessment using Doppler techniques and color flow imaging are common. Increased use of ultrasound is due to low equipment cost, portability, high safety, and minimal risk. Details of ultrasound are found in Chapter 16.

ducer, which records the returning sound beam. This mode of operation of an ultrasound device is called *pulse echo* imaging. The sound beam is swept over a range of angles (a sector) and the echoes from each line are recorded and used to compute an ultrasonic image in the shape of a sector (*sector scanning*).

Ultrasound is reflected strongly by interfaces, such as the surfaces and internal structures of abdominal organs. Because ultrasound is less harmful than ionizing radiation to a growing fetus, ultrasound imaging is preferred in obstetric patients (Fig. 1-9). An interface between tissue and air is highly echoic, and thus very little sound can penetrate from tissue into an air-filled cavity. Therefore, ultrasound imaging has less utility in the thorax where the air in the lungs presents a wall that the sound beam cannot penetrate. Similarly, an interface between tissue and bone is also highly echoic, thus making brain imaging, for example, impractical in most cases.

Doppler Ultrasound Imaging

Doppler imaging using ultrasound makes use of a phenomenon familiar to train enthusiasts. For the observer standing beside railroad tracks as a rapidly moving train goes by blowing its whistle, the pitch of the whistle is higher as the train approaches and becomes lower as the train passes by the observer and speeds off into the distance. The change in the pitch of the whistle, which is an apparent change in the frequency of the sound, is a result of the Doppler effect. The same phenomenon occurs at ultrasound frequencies, and the change in frequency (the Doppler shift) is used to measure the motion of blood or of the heart. Both the velocity and direction of blood flow can be measured, and color Doppler display usually shows blood flow in one direction as red and in the other direction as blue.

1.2 IMAGE PROPERTIES

Contrast

Contrast in an image is the difference in the gray scale of the image. A uniformly gray image has no contrast, whereas an image with vivid transitions between dark gray and light gray demonstrates high contrast. Each imaging modality generates contrast based on different physical parameters in the patient.

X-ray contrast (radiography, fluoroscopy, mammography, and CT) is produced by differences in tissue composition, which affect the local x-ray absorption coefficient, which in turn is dependent on the density (g/cm^3) and the effective atomic number. The energy of the x-ray beam (adjusted by the operator) also affects contrast in x-ray images. Because bone has a markedly different effective atomic number ($Z_{\text{eff}} \approx 13$) than soft tissue ($Z_{\text{eff}} \approx 7$), due to its high concentration of calcium ($Z = 20$) and phosphorus ($Z = 15$), bones produce high contrast on x-ray image modalities. The chest radiograph, which demonstrates the lung parenchyma with high tissue/airway contrast, is the most common radiographic procedure performed in the world (Fig. 1-2).

CT's contrast is enhanced over other radiographic imaging modalities due to its tomographic nature. The absence of out-of-slice structures in the CT image greatly improves its image contrast.

Nuclear medicine images (planar images, SPECT, and PET) are maps of the spatial distribution of radioisotopes in the patients. Thus, contrast in nuclear images depends on the tissue's ability to concentrate the radioactive material. The uptake of a radiopharmaceutical administered to the patient is dependent on the pharmacologic interaction of the agent with the body. PET and SPECT have much better contrast than planar nuclear imaging because, like CT, the images are not obscured by out-of-slice structures.

Contrast in MRI is related primarily to the proton density and to relaxation phenomena (i.e., how fast a group of protons gives up its absorbed energy). Proton density is influenced by the mass density (g/cm^3), so MRI can produce images that look somewhat like CT images. Proton density differs among tissue types, and in particular adipose tissues have a higher proportion of protons than other tissues, due to the high concentration of hydrogen in fat $[\text{CH}_3(\text{CH}_2)_n\text{COOH}]$. Two different relaxation mechanisms (spin/lattice and spin/spin) are present in tissue, and the dominance of one over the other can be manipulated by the timing of the radiofrequency (RF) pulse sequence and magnetic field variations in the MRI system. Through the clever application of different pulse sequences, blood flow can be detected using MRI techniques, giving rise to the field of MR angiography. Contrast mechanisms in MRI are complex, and thus provide for the flexibility and utility of MR as a diagnostic tool.

Contrast in ultrasound imaging is largely determined by the acoustic properties of the tissues being imaged. The difference between the *acoustic impedances* (tissue density $\times$ speed of sound in tissue) of two adjacent tissues or other substances affects the amplitude of the returning ultrasound signal. Hence, contrast is quite apparent at tissue interfaces where the differences in acoustic impedance are large. Thus, ultrasound images display unique information about patient anatomy not provided by other imaging modalities. Doppler ultrasound imaging shows the amplitude and direction of blood flow by analyzing the frequency shift in the reflected signal, and thus motion is the source of contrast.

Spatial Resolution

Just as each modality has different mechanisms for providing contrast, each modality also has different abilities to resolve fine detail in the patient. *Spatial resolution* refers to the ability to see small detail, and an imaging system has *higher* spatial resolution if it can demonstrate the presence of *smaller* objects in the image. The *limiting spatial resolution* is the size of the smallest object that an imaging system can resolve.

Table 1-1 lists the limiting spatial resolution of each imaging modality used in medical imaging. The wavelength of the energy used to probe the object is a fundamental limitation of the spatial resolution of an imaging modality. For example, optical microscopes cannot resolve objects smaller than the wavelengths of visible light, about 500 nm. The wavelength of x-rays depends on the x-ray energy, but even the longest x-ray wavelengths are tiny—about one ten-billionth of a meter. This is far from the actual resolution in x-ray imaging, but it does represent the theoretic limit on the spatial resolution using x-rays. In ultrasound imaging, the wavelength of sound is the fundamental limit of spatial resolution. At 3.5 MHz, the wavelength of sound in soft tissue is about 0.50 mm. At 10 MHz, the wavelength is 0.15 mm.

TABLE 1-1. THE LIMITING SPATIAL RESOLUTIONS OF VARIOUS MEDICAL IMAGING MODALITIES: THE RESOLUTION LEVELS ACHIEVED IN TYPICAL CLINICAL USAGE OF THE MODALITY

Modality	Δ (mm)	Comments
Screen film radiography	0.08	Limited by focal spot and detector resolution
Digital radiography	0.17	Limited by size of detector elements
Fluoroscopy	0.125	Limited by detector and focal spot
Screen film mammography	0.03	Highest resolution modality in radiology
Digital mammography	0.05–0.10	Limited by size of detector elements
Computed tomography	0.4	About $\frac{1}{2}$ -mm pixels
Nuclear medicine planar imaging	7	Spatial resolution degrades substantially with distance from detector
Single photon emission computed tomography	7	Spatial resolution worst toward the center of cross-sectional image slice
Positron emission tomography	5	Better spatial resolution than with the other nuclear imaging modalities
Magnetic resonance imaging	1.0	Resolution can improve at higher magnetic fields
Ultrasound imaging (5 MHz)	0.3	Limited by wavelength of sound

MRI poses a paradox to the wavelength imposed resolution rule—the wavelength of the RF waves used (at 1.5 T, 63 MHz) is 470 cm, but the spatial resolution of MRI is better than a millimeter. This is because the spatial distribution of the paths of RF energy is not used to form the actual image (contrary to ultrasound, light microscopy, and x-ray images). The RF energy is collected by a large antenna, and it carries the spatial information of a group of protons encoded in its frequency spectrum.

RADIATION AND THE ATOM

2.1 RADIATION

Radiation is energy that travels through space or matter. Two types of radiation used in diagnostic imaging are electromagnetic (EM) and particulate.

Electromagnetic Radiation

Visible light, radio waves, and x-rays are different types of EM radiation. EM radiation has no mass, is unaffected by either electrical or magnetic fields, and has a constant speed in a given medium. Although EM radiation propagates through matter, it does not require matter for its propagation. Its maximum speed (2.998×10^8 m/sec) occurs in a vacuum. In other media, its speed is a function of the transport characteristics of the medium. EM radiation travels in straight lines; however, its trajectory can be altered by interaction with matter. This interaction can occur either by *absorption* (removal of the radiation) or *scattering* (change in trajectory).

EM radiation is characterized by wavelength (λ), frequency (ν), and energy per photon (E). Categories of EM radiation (including radiant heat; radio, TV, and microwaves; infrared, visible, and ultraviolet light; and x- and gamma rays) comprise the electromagnetic spectrum (Fig. 2-1).

EM radiation used in diagnostic imaging include: (a) *gamma rays*, which emanate from within the nuclei of radioactive atoms and are used to image the distribution of radiopharmaceuticals; (b) *x-rays*, which are produced outside the nucleus and are used in radiography and computed tomography imaging; (c) *visible light*, which is produced in detecting x- and gamma rays and is used for the observation and interpretation of images; and (d) *radiofrequency* EM radiation in the FM region, which is used as the transmission and reception signal for magnetic resonance imaging (MRI).

There are two equally correct ways of describing EM radiation—as waves and as particle-like units of energy called *photons* or *quanta*. In some situations EM radiation behaves like waves and in other situations like particles.

Wave Characteristics

All waves (mechanical or electromagnetic) are characterized by their *amplitude* (maximal height), *wavelength* (λ), *frequency* (ν), and *period*. The amplitude is the intensity of the wave. The wavelength is the distance between any two identical

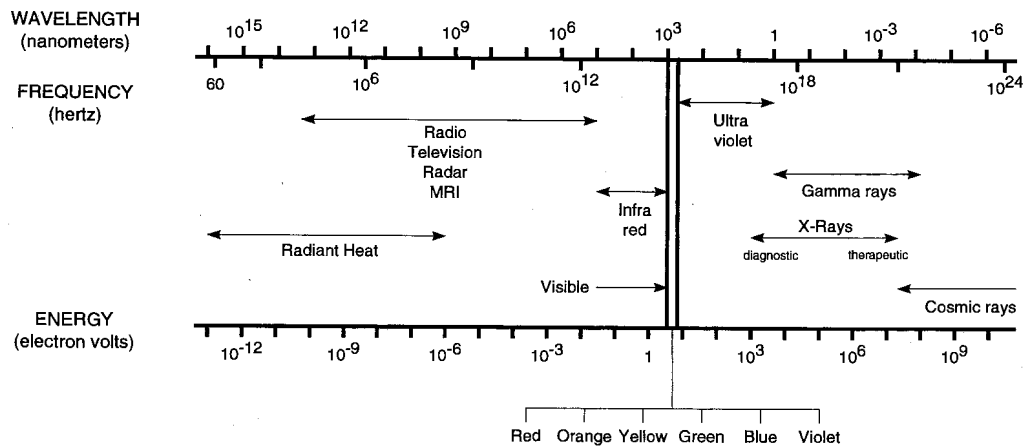


FIGURE 2-1. The electromagnetic spectrum.

points on adjacent cycles. The time required to complete one cycle of a wave (i.e., one λ) is the period. The number of periods that occur per second is the frequency (1/period). Phase is the temporal shift of one wave with respect to the other. Some of these quantities are depicted in Fig. 2-2. The speed (c), wavelength, and frequency of all waves are related by

$$c = \lambda \nu \quad [2-1]$$

Because the speed of EM radiation is essentially constant, its frequency and wavelength are inversely proportional. Wavelengths of x-rays and gamma rays are typically measured in *nanometers* (nm), where $1 \text{ nm} = 10^{-9} \text{ m}$. Frequency is expressed in *hertz* (Hz), where $1 \text{ Hz} = 1 \text{ cycle/sec} = 1 \text{ sec}^{-1}$.

EM radiation propagates as a pair of perpendicular electric and magnetic fields, as shown in Fig. 2-3.

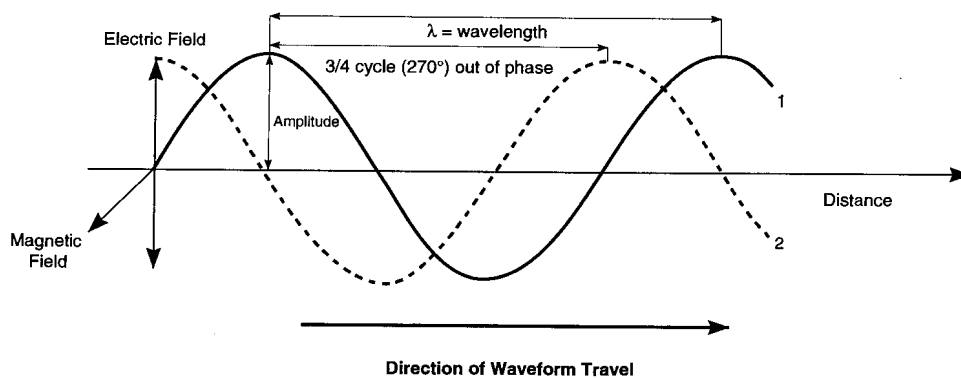


FIGURE 2-2. Characterization of waves.

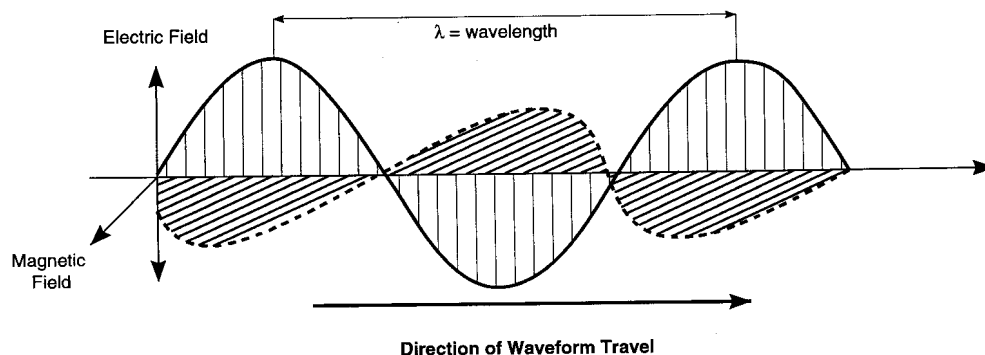


FIGURE 2-3. Electric and magnetic field components of electromagnetic radiation.

Problem: Find the frequency of blue light with a wavelength of 400 nm.

Solution: From Equation 2-1:

$$\nu = c/\lambda = [(3 \times 10^8 \text{ m/sec})/400 \text{ nm}](10^9 \text{ nm/m}) = 7.5 \times 10^{14} \text{ sec}^{-1} = 7.5 \times 10^{14} \text{ Hz}$$

Particle Characteristics

When interacting with matter, EM radiation can exhibit particle-like behavior. These particle-like bundles of energy are called *photons*. The energy of a photon is given by

$$E = h\nu = hc/\lambda \quad [2-2]$$

where h (Planck's constant) = $6.62 \times 10^{-34} \text{ J}\cdot\text{sec} = 4.13 \times 10^{-18} \text{ keV}\cdot\text{sec}$. When E is expressed in keV and λ in nanometers (nm):

$$E \text{ (keV)} = \frac{1.24}{\lambda \text{ (nm)}} \quad [2-3]$$

The energies of photons are commonly expressed in electron volts (eV). One electron volt is defined as the energy acquired by an electron as it traverses an electrical potential difference (voltage) of one volt in a vacuum. Multiples of the eV common to medical imaging are the keV (1,000 eV) and the MeV (1,000,000 eV).

Ionizing vs. Nonionizing Radiation

EM radiation of higher frequency than the near-ultraviolet region of the spectrum carries sufficient energy per photon to remove bound electrons from atomic shells, thus producing ionized atoms and molecules. Radiation in this portion of the spectrum (e.g., ultraviolet radiation, x-rays, and gamma rays) is called *ionizing radiation*. EM radiation with energy below the far-ultraviolet region (e.g., visible light, infrared, radio and TV broadcasts) is called *nonionizing radiation*. The threshold energy for ionization depends on the type of matter. For example, the minimum energies necessary to remove an electron (referred to as the *ionization potential*) from H_2O and C_6H_6 are 12.6 and 9.3 eV, respectively.

Particulate Radiation

The physical properties of the most important particulate radiations associated with medical imaging are listed in Table 2-1. Protons are found in the nuclei of all atoms. A proton has a single positive charge and is the nucleus of a hydrogen-1 atom. Electrons exist in atomic orbits. Electrons are also emitted by the nuclei of some radioactive atoms; in this case they are referred to as *beta-minus particles*. Beta-minus particles (β^-) are also referred to as *negatrons* or simply “beta particles.” Positrons are positively charged electrons (β^+), and are emitted from some nuclei during radioactive decay. A neutron is an uncharged nuclear particle that has a mass slightly greater than that of a proton. Neutrons are released by nuclear fission and are used for radionuclide production. An alpha particle (α^{2+}) consists of two protons and two neutrons; it thus has a +2 charge and is identical to the nucleus of a helium atom (${}^4\text{He}^{2+}$). Alpha particles are emitted by certain naturally occurring radioactive materials, such as uranium, thorium, and radium. Following such emissions the α^{2+} particle eventually acquires two electrons from the surrounding medium and becomes a neutral helium atom (${}^4\text{He}$).

Mass Energy Equivalence

Einstein’s theory of relativity states that mass and energy are interchangeable. In any reaction, the sum of the mass and energy must be conserved. In classical physics, there are two separate conservation laws, one for mass and one for energy. Einstein showed that these conservation laws are valid only for objects moving at low speeds. The speeds associated with some nuclear processes approach the speed of light. At these speeds, mass and energy are equivalent according to the expression

$$E = mc^2 \quad [2-4]$$

where E represents the energy equivalent to mass m at rest and c is the speed of light in a vacuum (2.998×10^8 m/sec). For example, the energy equivalent of an electron ($m = 9.109 \times 10^{-31}$ kg) is

$$E = mc^2$$

$$E = (9.109 \times 10^{-31} \text{ kg}) (2.998 \times 10^8 \text{ m/sec})^2$$

$$E = 8.187 \times 10^{-14} \text{ J}$$

$$E = (8.187 \times 10^{-14} \text{ J}) (1 \text{ MeV}/1.602 \times 10^{-13} \text{ J})$$

$$E = 0.511 \text{ MeV or } 511 \text{ keV}$$

TABLE 2-1. FUNDAMENTAL PROPERTIES OF PARTICULATE RADIATION

Particle	Symbol	Relative Charge	Mass (amu)	Approximate Energy Equivalent (MeV)
Alpha	α , ${}^4\text{He}^{2+}$	+2	4.0028	3727
Proton	p , ${}^1\text{H}^+$	+1	1.007593	938
Electron (beta minus)	e^- , β^-	-1	0.000548	0.511
Positron (beta plus)	e^+ , β^+	+1	0.000548	0.511
Neutron	n^0	0	1.008982	940

amu, atomic mass unit

The atomic mass unit (amu) is defined as $1/12$ th of the mass of an atom of ^{12}C . It can be shown that 1 amu is equivalent to 931 MeV of energy.

2.2 STRUCTURE OF THE ATOM

The atom is the smallest division of an element in which the chemical identity of the element is maintained. The atom is composed of an extremely dense positively charged nucleus containing protons and neutrons and an extranuclear cloud of light negatively charged electrons. In its nonionized state, an atom is electrically neutral because the number of protons equals the number of electrons. The radius of an atom is approximately 10^{-10} m, whereas that of the nucleus is only about 10^{-14} m. Thus, the atom is largely unoccupied space, in which the volume of the nucleus is only 10^{-12} (a millionth of a millionth) the volume of the atom. If the empty space in an atom could be removed, a cubic centimeter of protons would have a mass of approximately 4 million metric tons.

Electronic Structure

Electron Orbits and Electron Binding Energy

In the Bohr model of the atom (proposed by Niels Bohr in 1913) electrons orbit around a dense positively charged nucleus at fixed distances. Each electron occupies a discrete energy state in a given electron shell. These electron shells are assigned the letters *K*, *L*, *M*, *N*, ..., with *K* denoting the innermost shell. The shells are also assigned the *quantum numbers* 1, 2, 3, 4, ..., with the quantum number 1 designating the *K* shell. Each shell can contain a maximum number of electrons given by $(2n^2)$, where n is the quantum number of the shell. Thus, the *K* shell ($n = 1$) can only hold 2 electrons, the *L* shell ($n = 2$) can hold $2(2)^2$ or 8 electrons, and so on, as shown in Fig. 2-4.

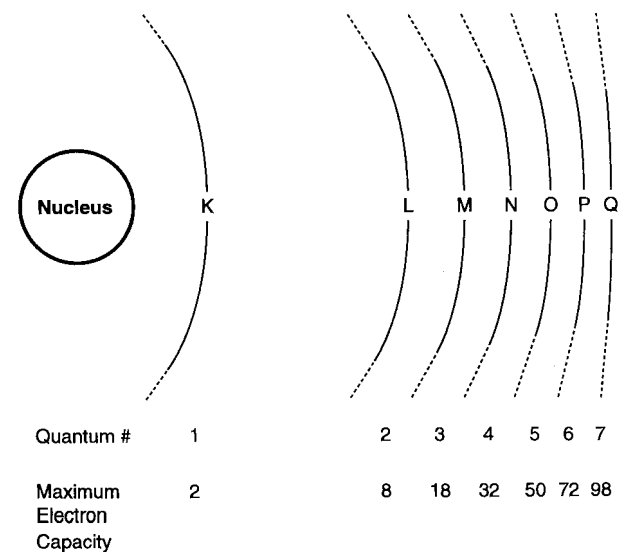


FIGURE 2-4. Electron shell designations and orbital filling rules.

The energy required to remove an electron completely from the atom is called its *binding energy*. By convention, binding energies are negative with increasing magnitude for electrons in shells closer to the nucleus. For an electron to become ionized, the energy transferred to it from a photon or particulate form of ionizing radiation must equal or exceed the magnitude of the electron's binding energy. The binding energy of electrons in a particular orbit increases with the number of protons in the nucleus (i.e., atomic number). In Fig. 2-5, binding energies are compared for electrons in hydrogen ($Z = 1$) and tungsten ($Z = 74$). If a free (unbound) electron is assumed to have a total energy of zero, the total energy of a bound electron is zero minus its binding energy. A K shell electron of tungsten is much more tightly bound ($-69,500$ eV) than the K shell electron of hydrogen (-13.5 eV). The energy required to move an electron from the innermost electron orbit (K shell) to the next orbit (L shell) is the difference between the binding energies of the two orbits (i.e., $E_{bK} - E_{bL}$ equals the transition energy).

Hydrogen:

$$13.5 \text{ eV} - 3.4 \text{ eV} = 10.1 \text{ eV}$$

Tungsten:

$$69,500 \text{ eV} - 11,000 \text{ eV} = 58,500 \text{ eV} (58.5 \text{ keV})$$

Advances in atomic physics and quantum mechanics have led to refinements of the Bohr model. According to contemporary views on atomic structure, orbital electrons are assigned probabilities for occupying any location within the atom. Nevertheless, the greatest probabilities are associated with Bohr's original atomic radii. The outer electron shell of an atom, the *valence shell*, determines the chemical properties of the element.

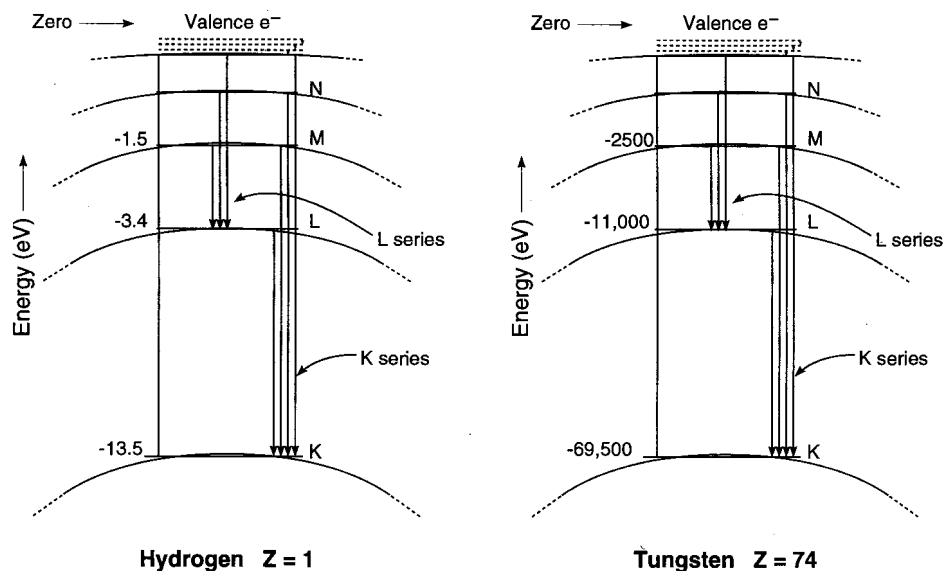


FIGURE 2-5. Energy-level diagrams for hydrogen and tungsten. Energies associated with various electron orbits (not drawn to scale) increase with Z and decrease with distance from the nucleus.

Radiation from Electron Transitions

When an electron is removed from its shell by an x- or gamma-ray photon or a charged particle interaction, a vacancy is created in that shell. This vacancy is usually filled by an electron from an outer shell, leaving a vacancy in the outer shell that in turn may be filled by an electron transition from a more distant shell. This series of transitions is called an *electron cascade*. The energy released by each transition is equal to the difference in binding energy between the original and final shells of the electron. This energy may be released by the atom as characteristic x-rays or Auger electrons.

Characteristic X-Rays

Electron transitions between atomic shells results in the emission of radiation in the visible, ultraviolet, and x-ray portions of the EM spectrum. The energy of this radiation is characteristic of each atom, since the electron binding energies depend on Z . Emissions from transitions exceeding 100 eV are called *characteristic* or *fluorescent* x-rays. Characteristic x-rays are named according to the orbital in which the vacancy occurred. For example, the radiation resulting from a vacancy in the K shell is called a K -characteristic x-ray, and the radiation resulting from a vacancy in the L shell is called an L -characteristic x-ray. If the vacancy in one shell is filled by the adjacent shell it is identified by a subscript alpha (e.g., $L \rightarrow K$ transition = K_α , $M \rightarrow L$ transition = L_α). If the electron vacancy is filled from a nonadjacent shell, the subscript beta is used (e.g., $M \rightarrow K$ transition = K_β). The energy of the characteristic x-ray ($E_{\text{Characteristic}}$) is the difference between the electron binding energies (E_b) of the respective shells:

$$E_{\text{Characteristic}} = E_b \text{ vacant shell} - E_b \text{ transition shell} \quad [2-5]$$

Thus, as illustrated in Fig. 2-6A, an M to K shell transition in tungsten would produce a K_β characteristic x-ray of

$$E(K_\beta) = E_{bK} - E_{bM}$$

$$E(K_\beta) = 69.5 \text{ keV} - 2.5 \text{ keV} = 67 \text{ keV}$$

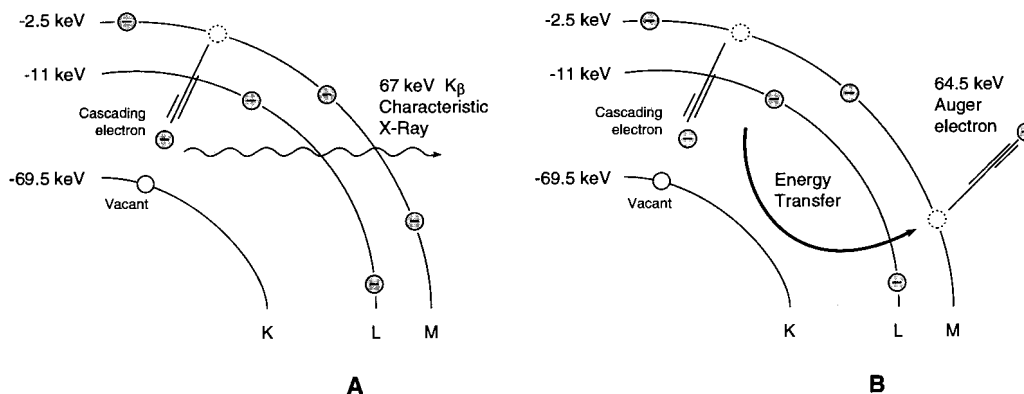


FIGURE 2-6. De-excitation of a tungsten atom. An electron transition filling a vacancy in an orbit closer to the nucleus will be accompanied by either the emission of characteristic radiation (**A**) or the emission of an Auger electron (**B**).

Auger Electrons And Fluorescent Yield

An electron cascade does not always result in the production of a characteristic x-ray. A competing process that predominates in low Z elements is *Auger electron emission*. In this case, the energy released is transferred to an orbital electron, typically in the same shell as the cascading electron (Fig. 2-6B). The ejected Auger electron possesses kinetic energy equal to the difference between the transition energy and the binding energy of the ejected electron.

The probability that the electron transition will result in the emission of a characteristic x-ray is called the *fluorescent yield* (ω). Thus $1 - \omega$ is the probability that the transition will result in the ejection of an Auger electron. Auger emission predominates in low Z elements and in electron transitions of the outer shells of heavy elements. The K -shell fluorescent yield is essentially zero ($\leq 1\%$) for elements $Z < 10$ (i.e., the elements comprising the majority of soft tissue), about 15% for calcium ($Z = 20$), about 65% for iodine ($Z = 53$), and approaches 80% for $Z > 60$.

The Atomic Nucleus

Composition of the Nucleus

The nucleus is composed of protons and neutrons (known collectively as *nucleons*). The number of protons in the nucleus is the *atomic number* (Z) and the total number of protons and neutrons (N) within the nucleus is the *mass number* (A). It is important not to confuse the mass number with the atomic mass, which is the actual mass of the atom. For example, the mass number of oxygen-16 is 16 (8 protons and 8 neutrons), whereas its atomic mass is 15.9949 amu. The notation specifying an atom with the chemical symbol X is A_ZX_N . In this notation, Z and X are redundant because the chemical symbol identifies the element and thus the number of protons. For example, the symbols H, He, and Li refer to atoms with $Z = 1$, 2, and 3, respectively. The number of neutrons is calculated as $N = A - Z$. For example, ${}^{131}_{53}\text{I}_{78}$ is usually written as ${}^{131}\text{I}$ or as I-131. The charge on an atom is indicated by a superscript to the right of the chemical symbol. For example, Ca^{+2} indicates that the calcium atom has lost two electrons.

Nuclear Forces and Energy Levels

There are two main forces that act in opposite directions on particles in the nucleus. The coulombic force between the protons is repulsive and is countered by the attractive force resulting from the exchange of pions (subnuclear particles) among all nucleons. These exchange forces, also called the *strong forces*, hold the nucleus together but operate only over very short nuclear distances ($< 10^{-14}$ m).

The nucleus has energy levels that are analogous to orbital electron shells, although often much higher in energy. The lowest energy state is called the *ground state* of an atom. Nuclei with energy in excess of the ground state are said to be in an *excited state*. The average lifetimes of excited states range from 10^{-16} seconds to more than 100 years. Excited states that exist longer than 10^{-12} seconds are referred to as *metastable* or *isomeric states*. These excited nuclei are denoted by the letter m after the mass number of the atom (e.g., Tc-99m).

Classification of Nuclides

Species of atoms characterized by the number of protons, neutrons, and the energy content of the atomic nucleus are called *nuclides*. Isotopes, isobars, isotones, and isomers are families of nuclides that share specific properties (Table 2-2). An easy way to remember these relationships is to associate the *p* in *isotopes* with the same number of protons, the *a* in *isobars* with the same atomic mass number, the *n* in *isotones* with the same number of neutrons, and the *e* in *isomer* with the different nuclear energy states.

Nuclear Stability

Only certain combinations of neutrons and protons in the nucleus are stable. On a plot of *Z* versus *N* these nuclides fall along a “line of stability” for which the *N/Z* ratio is approximately 1 for low *Z* nuclides and approximately 1.5 for high *Z* nuclides, as shown in Fig. 2-7. A higher neutron-to-proton ratio is required in heavy elements to offset the coulombic repulsive forces between protons by providing increased separation of protons. Nuclei with an odd number of neutrons and an odd number of protons tend to be unstable, whereas nuclei with an even number of neutrons and an even number protons more frequently are stable. The number of stable nuclides identified for different combinations of neutrons and protons is shown in Table 2-3. Nuclides with an odd number of nucleons are capable of producing a nuclear magnetic resonance (NMR) signal, as described in Chapter 14.

Radioactivity

Combinations of neutrons and protons that are not stable do exist but over time they will permute to nuclei that are stable. Two kinds of instability are neutron excess and neutron deficiency (i.e., proton excess). Such nuclei have excess internal energy compared with a stable arrangement of neutrons and protons. They achieve stability by the conversion of a neutron to a proton or vice versa, and these events are accompanied by the emission of energy. The energy emissions include particulate and electromagnetic radiations. Nuclides that decay (i.e., transform) to more stable nuclei are said to be *radioactive*, and the transformation process itself is called *radioactive decay* or radioactive disintegration. There are several types of radioactive decay and these are discussed in detail in Chapter 18. A nucleus may undergo sev-

TABLE 2-2. NUCLEAR FAMILIES: ISOTOPES, ISOBARS, ISOTONES, AND ISOMERS

Family	Nuclides with Same	Example
Isotopes	Atomic number (<i>Z</i>)	I-131 and I-125: <i>Z</i> = 53
Isobars	Mass number (<i>A</i>)	Mo-99 and Tc-99: <i>A</i> = 99
Isotones	Number of neutrons (<i>A</i> – <i>Z</i>)	⁵³ I-131: 131 – 53 = 78 ⁵⁴ Xe-132: 132 – 54 = 78
Isomers	Atomic and mass numbers but different energy states	Tc-99m and Tc-99: <i>Z</i> = 43 <i>A</i> = 99 Energy of Tc-99m > Tc-99: Δ <i>E</i> = 142 keV

Note: See text for description of the italicized letters in the nuclear family terms.

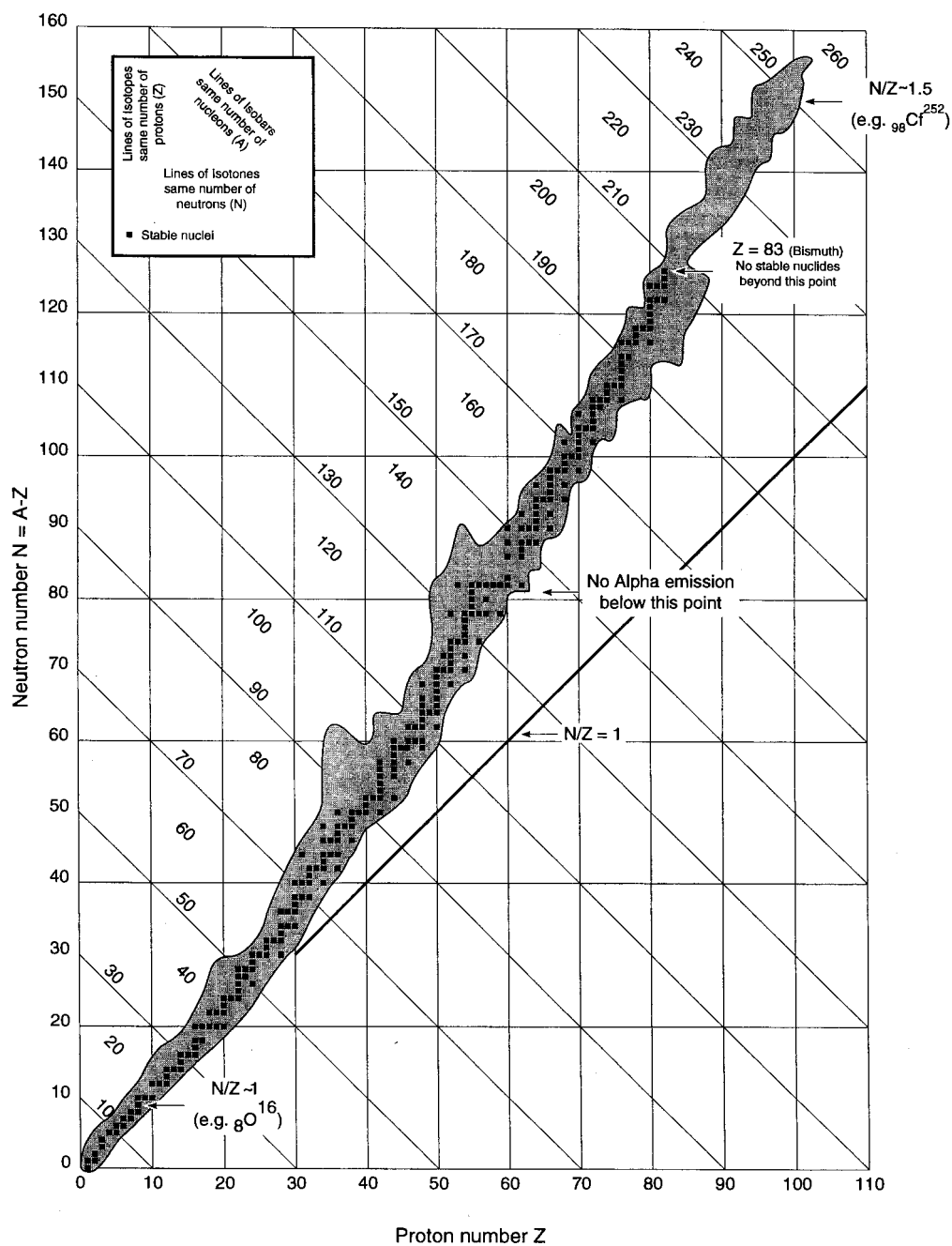


FIGURE 2-7. Nuclide line of stability. The shaded area represents the range of known nuclides. The stable nuclides are indicated by small black squares, whereas all other locations in the shaded area represent radioactive (i.e., unstable) nuclides. Note that all nuclides with $Z > 83$ are radioactive.

TABLE 2-3. DISTRIBUTION OF STABLE NUCLIDES AS A FUNCTION OF NEUTRON AND PROTON NUMBER

Number of Protons (Z)	Number of Neutrons (N)	Number of Stable Nuclides
Even	Even	165
Even	Odd	57 (NMR signal)
Odd	Even	53 (NMR signal)
Odd	Odd	4
Total		279

NMR, nuclear magnetic resonance.

eral decays before a stable configuration is achieved. These “decay chains” are often found in nature. For example, the decay of uranium 238 (U-238) is followed by 14 successive decays before the stable nuclide, lead 206 (Pb-206), is formed. The radionuclide at the beginning of a particular decay sequence is referred to as the *parent*, and the nuclide produced by the decay of the parent is referred to as the *daughter*. The daughter may be either stable or radioactive.

Gamma Rays

Radioactive decay often results in the formation of a daughter nucleus in an excited state. The EM radiation emitted from the nucleus as the excited state decays to a lower (more stable) energy state is called a *gamma ray*. This energy transition is analogous to the emission of characteristic x-rays following electron transition. However, gamma rays (by definition), emanate from the nucleus. Because the spacing of the energy states within the nucleus is often considerably larger than those associated with electron transitions, gamma rays are often much more energetic than characteristic x-rays. When this nuclear de-excitation process takes place in a metastable isomer (e.g., Tc-99m), it is called *isomeric transition*. In isomeric transition, the nuclear energy state is reduced with no change in A or Z.

Internal Conversion Electrons

Nuclear de-excitation does not always result in the emission of a gamma ray. An alternative form of de-excitation is *internal conversion*, in which the de-excitation energy is completely transferred to an orbital (typically K, L, or M) electron. The conversion electron is immediately ejected from the atom, with kinetic energy equal to that of the gamma ray less the electron binding energy. The vacancy produced by the ejection of the conversion electron will be filled by an electron cascade as described previously.

Nuclear Binding Energy and Mass Defect

The energy required to separate an atom into its constituent parts is the *atomic binding energy*. It is the sum of the electron binding energy and the nuclear binding energy. The *nuclear binding energy* is the energy necessary to disassociate a nucleus into its constituent parts and is the result of the strong attractive forces between nucleons. Compared with the nuclear binding energy, the electron binding energy is negligible. When two subatomic particles approach each other under the influence of this nuclear strong force, their total energy decreases and the lost energy is emitted in the form of radiation. Thus, the total energy of the bound particles is less than that of the separated free particles.

The binding energy can be calculated by subtracting the mass of the atom from the total mass of its constituent protons, neutrons, and electrons; this mass difference is called the *mass defect*. For example, the mass of an N-14 atom, which is composed of 7 electrons, 7 protons, and 7 neutrons, is 14.00307 amu. The total mass of its constituent particles in the unbound state is

$$\begin{aligned}\text{mass of 7 protons} &= 7(1.00727 \text{ amu}) = 7.05089 \text{ amu} \\ \text{mass of 7 neutrons} &= 7(1.00866 \text{ amu}) = 7.06062 \text{ amu} \\ \text{mass of 7 electrons} &= 7(0.00055 \text{ amu}) = \underline{0.00385 \text{ amu}} \\ \text{mass of component particles of N-14} &= 14.11536 \text{ amu}\end{aligned}$$

Thus, the mass defect of the N-14 atom, the difference between the mass of its constituent particles and its atomic mass, is $14.11536 \text{ amu} - 14.00307 \text{ amu} = 0.11229 \text{ amu}$. According to the formula for mass energy equivalence (Eq. 2-4), this mass defect is equal to $(0.11229 \text{ amu})(931 \text{ MeV/amu}) = 104.5 \text{ MeV}$.

An extremely important observation is made by carrying the binding energy calculation a bit further. The total binding energy of the nucleus may be divided by the mass number A to obtain the average binding energy per nucleon. Figure 2-8 shows the average binding energy per nucleon for stable nuclides as a function of mass number. The fact that this curve reaches its maximum near the middle elements and decreases at either end predicts that the energy contained in matter can be released. The two processes by which this can occur are called *nuclear fission* and *nuclear fusion*.

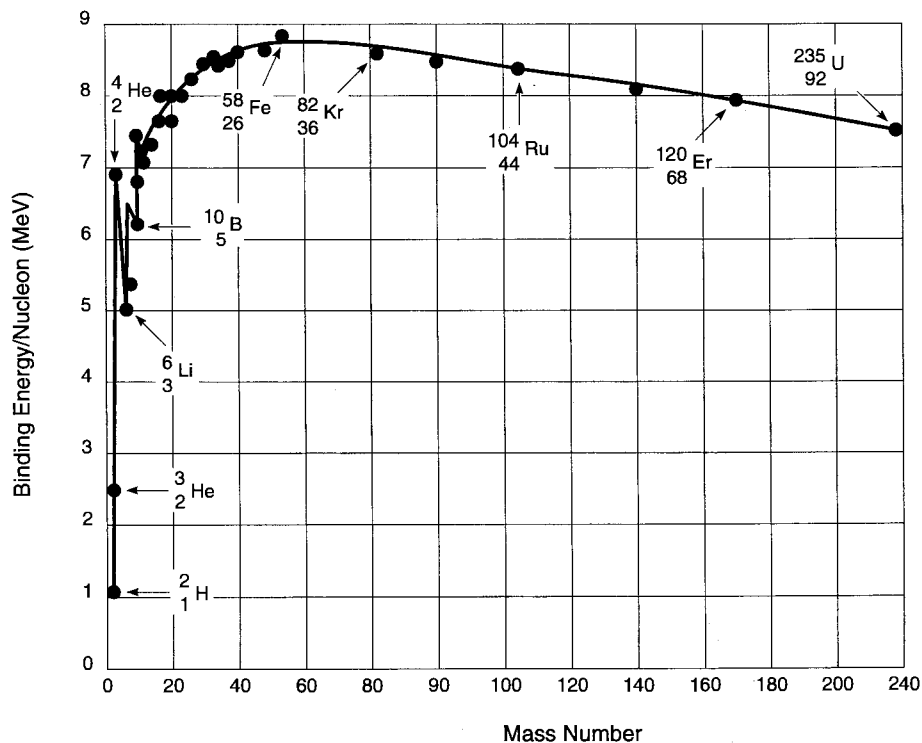


FIGURE 2-8. Average binding energy per nucleon.

Nuclear Fission and Fusion

During nuclear fission, a nucleus with a large atomic mass splits into two usually unequal parts called *fission fragments*, each with an average binding energy per nucleon greater than that of the original nucleus. In this reaction, the total nuclear binding energy increases. The change in the nuclear binding energy is released as electromagnetic radiation and as kinetic energy of particulate radiation. Fission usually releases energetic neutrons in addition to the heavier fission fragments. For example, fission of uranium 236 (U 236) may result in Sn 131 and Mo 102 fission fragments, and three free neutrons. This reaction results in a mass defect equivalent to approximately 200 MeV, which is released as EM radiation and kinetic energy of the fission fragments and neutrons. The probability of fission increases with neutron flux and this fact is used to produce energy in nuclear power plants and in the design of “atom” bombs. The use of nuclear fission for radionuclide production is discussed in Chapter 19.

Energy is also released from the combining or *fusion* of light atomic nuclei. For example, the fusion of deuterium (H-2) and helium-3 (He-3) results in the production of helium-4 and a proton. As can be seen in Fig. 2-8, the helium-4 atom has a much higher average binding energy per nucleon than either helium-3 or deuterium. The energy associated with the mass defect of this reaction is about 18.3 MeV. The nuclei involved in fusion reactions require extremely high kinetic energies in order to overcome coulombic repulsive forces. Fusion is self-sustaining in stars, where huge gravitational forces result in enormous temperatures. The so-called *H-bomb* is a fusion device that uses the detonation of an atom bomb (a fission device) to generate the temperature and pressure necessary for fusion. H-bombs are also referred to as “thermonuclear” weapons.

SUGGESTED READING

Evans RD. *The atomic nucleus*. Malabar, FL: Robert E. Krieger, 1982.

INTERACTION OF RADIATION WITH MATTER

3.1 PARTICLE INTERACTIONS

Particles of ionizing radiation include charged particles, such as alpha particles (α^{+2}), protons (p^{+}), electrons (e^{-}), beta particles (β^{-}), and positrons (e^{+} or β^{+}), and uncharged particles, such as neutrons. The behavior of heavy charged particles (e.g., alpha particles and protons) is different from that of lighter charged particles such as electrons and positrons.

Excitation, Ionization, and Radiative Losses

Energetic charged particles all interact with matter by electrical (i.e., coulombic) forces and lose kinetic energy via *excitation*, *ionization*, and *radiative losses*. Excitation and ionization occur when charged particles lose energy by interacting with orbital electrons. Excitation is the transfer of some of the incident particle's energy to electrons in the absorbing material, promoting them to electron orbitals farther from the nucleus (i.e., higher energy levels). In excitation, the energy transferred to an electron does not exceed its binding energy. Following excitation, the electron will return to a lower energy level, with the emission of the excitation energy in the form of electromagnetic radiation or Auger electrons. This process is referred to as *de-excitation* (Fig. 3-1A). If the transferred energy exceeds the binding energy of the electron, ionization occurs, whereby the electron is ejected from the atom (Fig. 3-1B). The result of ionization is an *ion pair* consisting of the ejected electron and the positively charged atom. Sometimes the ejected electrons possess sufficient energy to produce further ionizations called *secondary ionization*. These electrons are called *delta rays*.

Approximately 70% of charged particle energy deposition leads to nonionizing excitation. So while the smallest binding energies for electrons in carbon, nitrogen, and oxygen are less than approximately 10 eV, the average energy deposited per ion pair produced in air (mostly nitrogen and oxygen) or soft tissue (mostly hydrogen, carbon, and oxygen) is approximately 34 eV. The energy difference (approximately 24 eV) is the result of the excitation process.

Specific Ionization

The number of primary and secondary ion pairs produced per unit length of the charged particle's path is called the *specific ionization*, expressed in ion pairs (IP)/mm.

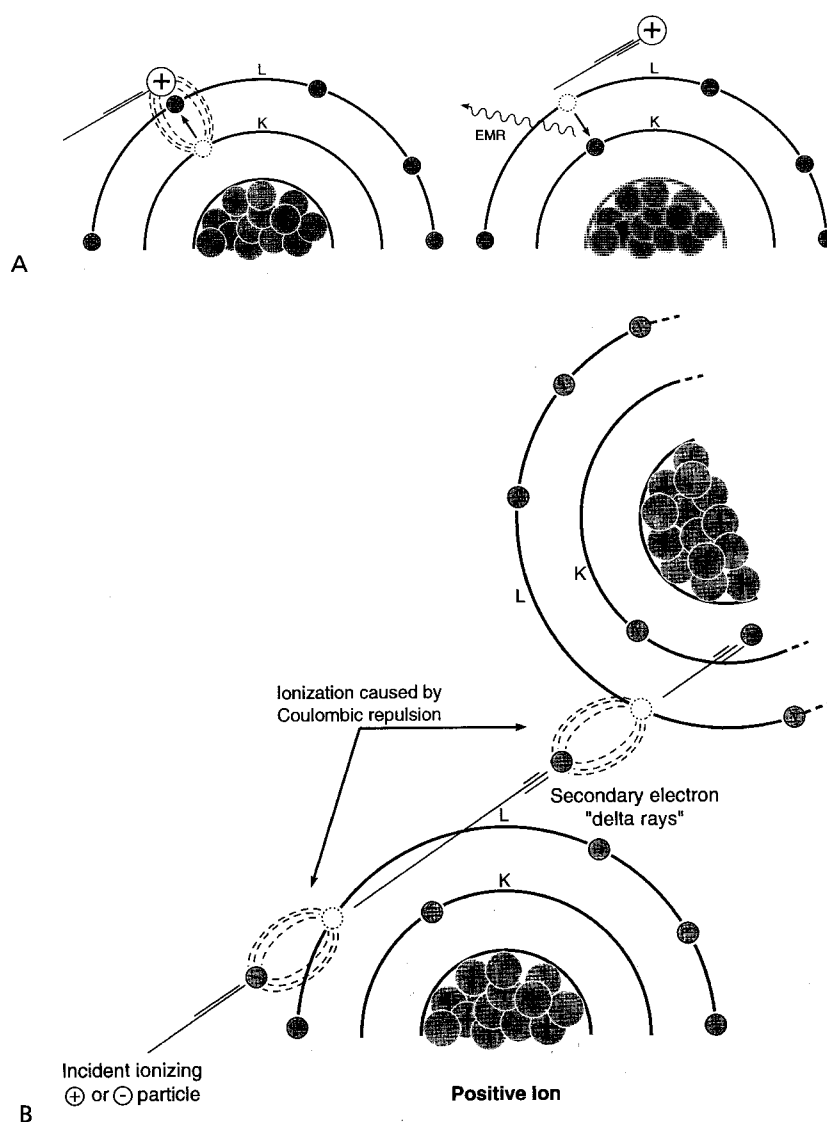


FIGURE 3-1. A: Excitation (*left*) and de-excitation (*right*) with the subsequent release of electromagnetic radiation. **B:** Ionization and the production of delta rays.

Specific ionization increases with the electrical charge of the particle and decreases with incident particle velocity. A larger charge produces a greater coulombic field; as the particle loses energy, it slows down, allowing the coulombic field to interact at a given location for a longer period of time. The specific ionization of an alpha particle can be as high as $\sim 7,000$ IP/mm in air. The specific ionization as a function of the particle's path is shown for a 7.69 MeV alpha particle from polonium-214 in air (Fig. 3-2). As the alpha particle slows, the specific ionization increases to a maximum (called the *Bragg peak*), beyond which it decreases rapidly as the alpha particle picks up electrons and

becomes electrically neutral, thus losing its capacity for further ionization. The large Bragg peak associated with heavy charged particles has medical applications in radiation therapy. By adjusting the kinetic energy of heavy charged particles, a large radiation dose can be delivered at a particular depth and over a fairly narrow range of tissue containing a lesion. On either side of the Bragg peak, the dose to tissue is substantially lower. Heavy particle accelerators are used at some medical facilities to provide this treatment in lieu of surgical excision or conventional radiation therapy. Compared to heavy charged particles, the specific ionization of electrons is much lower (in the range of 50 to 100 IP/cm of air).

Charged Particle Tracks

Another important distinction between heavy charged particles and electrons is their paths in matter. Electrons follow tortuous paths in matter as the result of multiple scattering events caused by coulombic deflections (repulsion and/or attraction). The sparse nonuniform ionization track of an electron is shown in Fig. 3-3A. On the other hand, the larger mass of a heavy charged particle results in a dense and usually linear ionization track (Fig. 3-3B). The *path length* of a particle is defined as the actual distance the particle travels. The *range* of a particle is defined as the actual depth of penetration of the particle in matter. As illustrated in Fig. 3-3, the path length of the electron almost always exceeds its range, whereas the straight ionization track of a heavy charged particle results in the path and range being nearly equal.

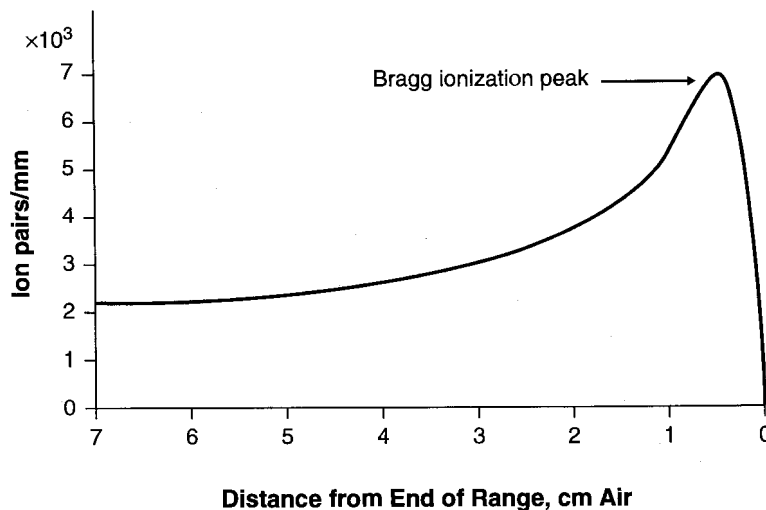


FIGURE 3-2. Specific ionization (ion pairs/mm) as a function of distance from the end of range in air for a 7.69-MeV alpha particle from polonium 214 (Po 214). Rapid increase in specific ionization reaches a maximum (Bragg peak) and then drops off sharply as the particle kinetic energy is exhausted and the charged particle is neutralized.

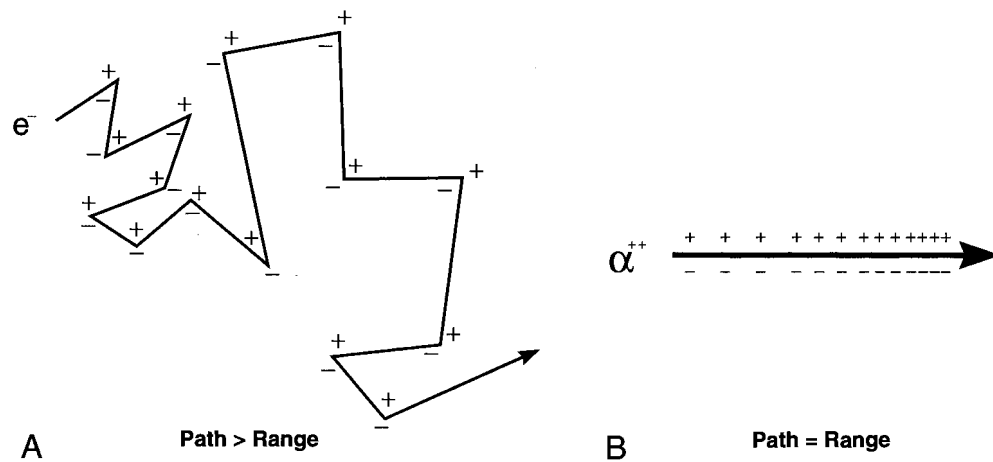


FIGURE 3-3. A: Electron scattering results in the path length of the electron being greater than its range. **B:** Heavily charged particles, like alpha particles, produce a dense nearly linear ionization track, resulting in the path and range being essentially equal.

Linear Energy Transfer

The amount of energy deposited per unit path length is called the *linear energy transfer* (LET) and is usually expressed in units of eV/cm. The LET of a charged particle is proportional to the square of the charge and inversely proportional to the particle's kinetic energy (i.e., LET proportional to Q^2/E_k). LET is the product of specific ionization (IP/cm) and the average energy deposited per ion pair (eV/IP). The LET of a particular type of radiation describes the energy deposition density, which largely determines the biologic consequence of radiation exposure. In general, "high LET" radiations (alpha particles, protons, etc.) are much more damaging to tissue than "low LET" radiations, which include electrons (e^- , β^- , and β^+) and ionizing electromagnetic radiation (gamma and x-rays, whose interactions set electrons into motion).

Scattering

Scattering refers to an interaction resulting in the deflection of a particle or photon from its original trajectory. A scattering event in which the total kinetic energy of the colliding particles is unchanged is called *elastic*. Billiard ball collisions, for example, are elastic (disregarding frictional losses). When scattering occurs with a loss of kinetic energy (i.e., the total kinetic energy of the scattered particles is less than that of the particles before the interaction), the interaction is said to be *inelastic*. For example, the process of ionization can be considered an elastic interaction if the binding energy of the electron is negligible compared to the kinetic energy of the incident electron (i.e., the kinetic energy of the ejected electron is equal to the kinetic energy lost by the incident electron). If the binding energy that must be overcome to ionize the atom is significant (i.e., the kinetic energy of the ejected electron is less than the kinetic energy lost by the incident electron), the process is said to be inelastic.

Radiative Interactions—Bremsstrahlung

Electrons can undergo inelastic interactions with atomic nuclei in which the path of the electron is deflected by the positively charged nucleus, with a loss of kinetic energy. This energy is instantaneously emitted as ionizing electromagnetic radiation (x-rays). The energy of the x-ray is equal to the energy lost by the electron, as required by the conservation of energy.

The radiation emission accompanying electron deceleration is called *bremsstrahlung*, a German word meaning “braking radiation” (Fig. 3-4). The deceleration of the high-speed electrons in an x-ray tube produces the bremsstrahlung x-rays used in diagnostic imaging.

When the kinetic energy of the electron is low, the bremsstrahlung photons are emitted predominantly between the angles of 60 and 90 degrees relative to the incident electron trajectory. At higher electron kinetic energies, the x-rays tend to be emitted in the forward direction. The probability of bremsstrahlung emission per atom is proportional to Z^2 of the absorber. Energy emission via bremsstrahlung varies inversely with the square of the mass of the incident particle. Therefore, protons and alpha particles will produce less than one-millionth the amount of bremsstrahlung radiation as electrons of the same energy. The ratio of electron energy loss by bremsstrahlung production to that lost by excitation and ionization can be calculated from Equation 3-1:

$$\frac{\text{Bremsstrahlung Radiation}}{\text{Excitation and Ionization}} = \frac{E_k Z}{820} \quad [3-1]$$

where E_k is the kinetic energy of the incident electron in MeV, and Z is the atomic number of the absorber. Bremsstrahlung x-ray production accounts for approximately 1% of the energy loss when electrons are accelerated to an energy of 100 keV and collide with a tungsten ($Z = 74$) target in an x-ray tube.

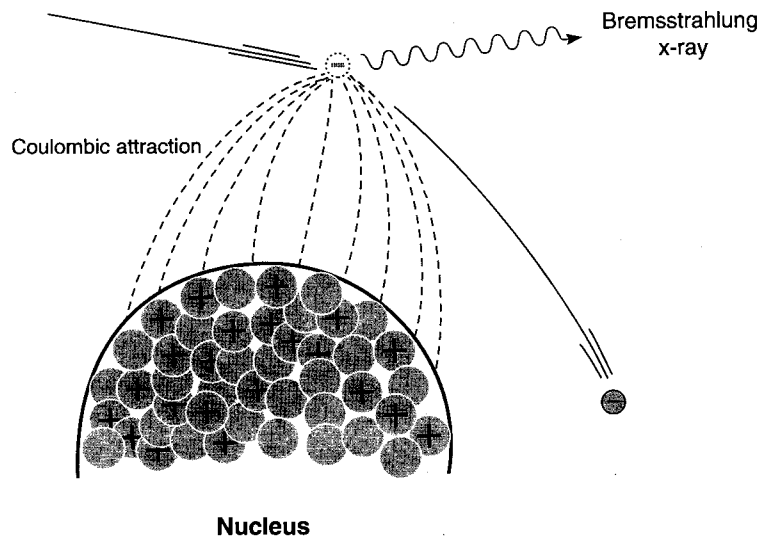


FIGURE 3-4. Radiative energy loss via bremsstrahlung (braking radiation).

The energy of a bremsstrahlung x-ray photon can be any value up to and including the entire kinetic energy of the deflected electron. Thus, when multiple electrons undergo bremsstrahlung interactions, the result is a continuous spectrum of x-ray energies. This radiative energy loss is responsible for the majority of the x-rays produced by x-ray tubes and is discussed in greater detail in Chapter 5.

Positron Annihilation

The fate of positrons (e^+ or β^+) is unlike that of negatively charged electrons (e^- and β^-) that ultimately become bound to atoms. As mentioned above, all electrons (positively and negatively charged) lose their kinetic energy by excitation, ionization, and radiative interactions. When a positron (a form of antimatter) comes to rest, it interacts with a negatively charged electron, resulting in the annihilation of the electron-positron pair, and the complete conversion of their rest mass to energy in the form of two oppositely directed 0.511 MeV *annihilation photons*. This process occurs following radionuclide decay by positron emission (see Chapter 18). Positrons are important in imaging of positron emitting radiopharmaceuticals in which the resultant annihilation photon pairs emitted from the patient are detected by positron emission tomography (PET) scanners (see Chapter 22).

Neutron Interactions

Neutrons are uncharged particles. Thus, they do not interact with electrons and therefore do not directly cause excitation and ionization. They do, however, interact with atomic nuclei, sometimes liberating charged particles or nuclear fragments that can directly cause excitation and ionization. Neutrons often interact with light atomic nuclei (e.g., H, C, O) via scattering in “billiard ball”-like collisions, producing recoil nuclei that lose their energy via excitation and ionization (Fig. 3-5). In tissue, energetic neutrons interact primarily with the hydrogen in water, producing recoil protons (hydrogen nuclei). Neutrons may also be captured by atomic nuclei. In some cases the neutron is reemitted; in other cases the neutron is retained, converting the atom to a different nuclide. In the latter case, the binding energy may be emitted via spontaneous gamma-ray emission:

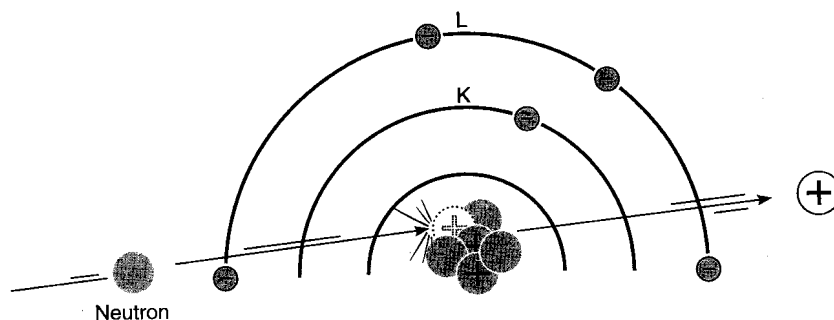


FIGURE 3-5. Collisional losses (uncharged particle interaction).

The nuclide produced by neutron absorption may be stable or radioactive. Neutron interactions are described in greater detail in Chapter 19.

3.2 X- AND GAMMA-RAY INTERACTIONS

When traversing matter, photons will penetrate, scatter, or be absorbed. There are four major types of interactions of x- and gamma-ray photons with matter, the first three of which play a role in diagnostic radiology and nuclear medicine: (a) Rayleigh scattering, (b) Compton scattering, (c) photoelectric absorption, and (d) pair production.

Rayleigh Scattering

In Rayleigh scattering, the incident photon interacts with and excites the *total atom*, as opposed to individual electrons as in Compton scattering or the photoelectric effect (discussed later). This interaction occurs mainly with very low energy diagnostic x-rays, as used in mammography (15 to 30 keV). During the Rayleigh scattering event, the electric field of the incident photon's electromagnetic wave expends energy, causing all of the electrons in the scattering atom to oscillate in phase. The atom's electron cloud immediately radiates this energy, emitting a photon of the same energy but in a slightly different direction (Fig. 3-6). In this interaction, electrons are not ejected and thus ionization does not occur. In general, the scattering angle increases as the x-ray energy decreases. In medical imaging, detection of the scattered x-ray will have a deleterious effect on image quality. However, this type of interaction has a low prob-

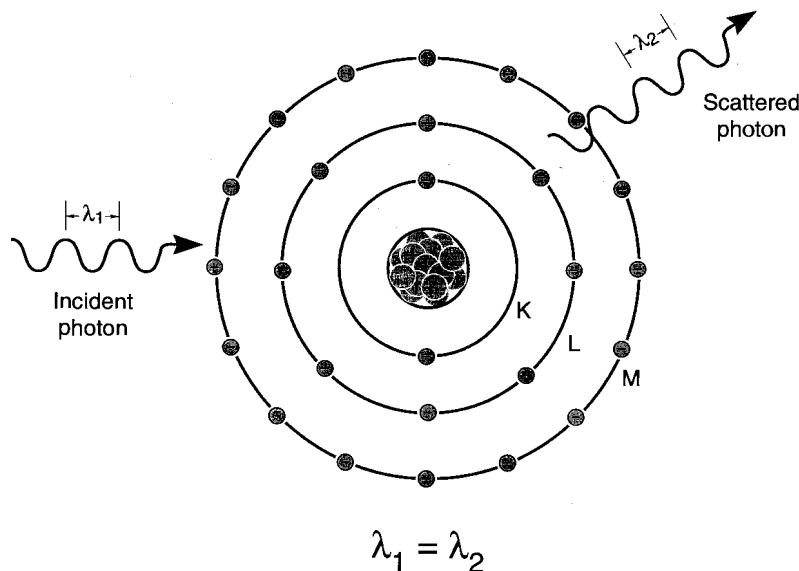


FIGURE 3-6. Rayleigh scattering. Diagram shows the incident photon λ_1 interacts with an atom and the scattered photon λ_2 is being emitted with approximately the same wavelength and energy. Rayleigh scattered photons are typically emitted in the forward direction fairly close to the trajectory of the incident photon. K, L, and M are electron shells.

ability of occurrence in the diagnostic energy range. In soft tissue, Rayleigh scattering accounts for less than 5% of x-ray interactions above 70 keV and at most only accounts for 12% of interactions at approximately 30 keV. Rayleigh interactions are also referred to as “coherent” or “classical” scattering.

Compton Scattering

Compton scattering (also called inelastic or nonclassical scattering) is the predominant interaction of x-ray and gamma-ray photons in the diagnostic energy range with soft tissue. In fact, Compton scattering not only predominates in the diagnostic energy range above 26 keV in soft tissue, but continues to predominate well beyond diagnostic energies to approximately 30 MeV. This interaction is most likely to occur between photons and outer (“valence”) shell electrons (Fig. 3-7). The electron is ejected from the atom, and the photon is scattered with some reduction in energy. As with all types of interactions, both energy and momentum must be conserved. Thus the energy of the incident photon (E_0) is equal to the sum of the energy of the scattered photon (E_{sc}) and the kinetic energy of the ejected electron (E_{e-}), as shown in Equation 3-2. The binding energy of the electron that was ejected is comparatively small and can be ignored.

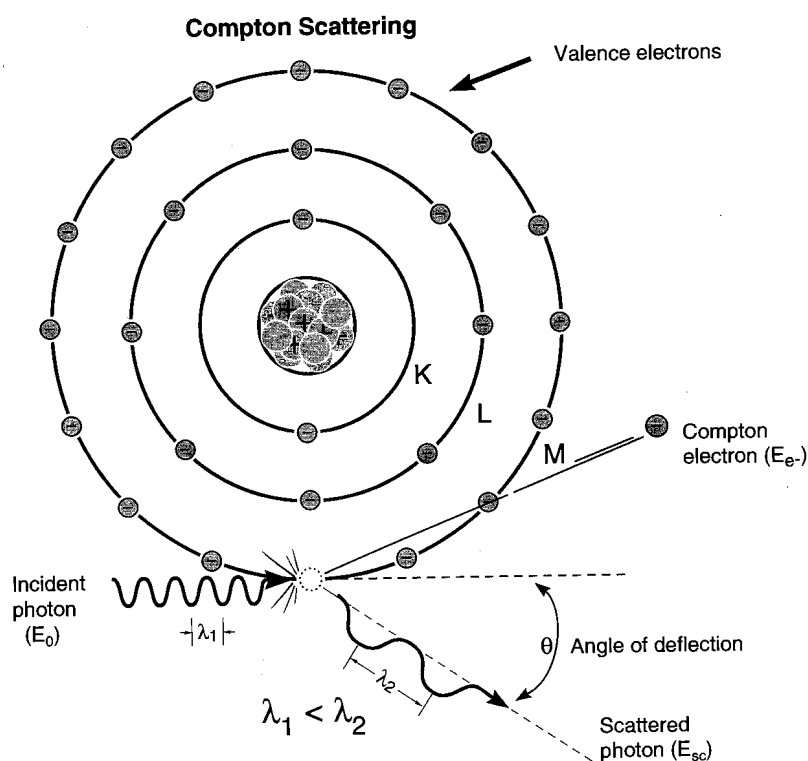


FIGURE 3-7. Compton scattering. Diagram shows the incident photon with energy E_0 , interacting with a valence shell electron that results in the ejection of the Compton electron (E_{e-}) and the simultaneous emission of a Compton scattered photon E_{sc} emerging at an angle θ relative to the trajectory of the incident photon. K, L, and M are electron shells.

$$E_o = E_{sc} + E_{e-} \quad [3-2]$$

Compton scattering results in the ionization of the atom and a division of the incident photon energy between the scattered photon and ejected electron. The ejected electron will lose its kinetic energy via excitation and ionization of atoms in the surrounding material. The Compton scattered photon may traverse the medium without interaction or may undergo subsequent interactions such as Compton scattering, photoelectric absorption (to be discussed shortly), or Rayleigh scattering.

The energy of the scattered photon can be calculated from the energy of the incident photon and the angle of the scattered photon (with respect to the incident trajectory):

$$E_{sc} = \frac{E_o}{1 + \frac{E_o}{511 \text{ keV}} (1 - \cos\theta)} \quad [3-3]$$

where E_{sc} = the energy of the scattered photon,
 E_o = the incident photon energy, and
 θ = the angle of the scattered photon.

As the incident photon energy increases, both scattered photons and electrons are scattered more toward the forward direction (Fig. 3-8). In x-ray transmission imaging, these photons are much more likely to be detected by the image receptor, thus reducing image contrast. In addition, for a given scattering angle, the fraction of energy transferred to the scattered photon decreases with increasing incident photon energy. Thus, for higher energy incident photons, the majority of the energy is transferred to the scattered electron. For example, for a 60-degree scattering angle, the scattered photon energy (E_{sc}) is 90% of the incident photon energy (E_o) at 100 keV, but only 17% at 5 MeV. When Compton scattering does occur at the lower x-ray energies used in diagnostic imaging (18 to 150 keV), the majority of the incident photon energy is transferred to the scattered photon, which, if detected by the image receptor, contributes to image degradation by reducing the primary photon

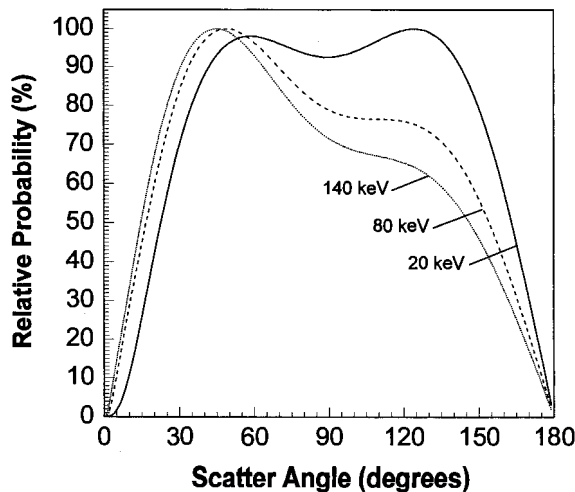


FIGURE 3-8. Graph illustrates relative Compton scatter probability as a function of scattering angle for 20-, 80-, and 140-keV photons in tissue. Each curve is normalized to 100%. (From Bushberg JT. The AAPM/RSNA physics tutorial for residents. X-ray interactions. *RadioGraphics* 1998;18:457–468, with permission.)

attenuation differences of the tissues. For example, following the Compton interaction of an 80-keV photon, the minimum energy of the scattered photon is 61 keV. Thus, even with maximal energy loss, the scattered photons have relatively high energies and tissue penetrability.

The laws of conservation of energy and momentum place limits on both scattering angle and energy transfer. For example, the maximal energy transfer to the Compton electron (and thus the maximum reduction in incident photon energy) occurs with a 180-degree photon backscatter. In fact, the maximal energy of the scattered photon is limited to 511 keV at 90 degrees scattering and to 255 keV for a 180-degree scattering (backscatter) event. These limits on scattered photon energy hold even for extremely high-energy photons (e.g., therapeutic energy range). The scattering angle of the ejected electron cannot exceed 90 degrees, whereas that of the scattered photon can be any value including a 180-degree backscatter. In contrast to the scattered photon, the energy of the ejected electron is usually absorbed near the scattering site.

The incident photon energy must be substantially greater than the electron's binding energy before a Compton interaction is likely to take place. Thus, the probability of a Compton interaction increases, compared to Rayleigh scattering or photoelectric absorption, as the incident photon energy increases. The probability of Compton interaction also depends on the electron density (number of electrons/g $\times$ density). With the exception of hydrogen, the total number of electrons/g is fairly constant in tissue; thus, the probability of Compton scattering per unit mass is nearly independent of Z , and the probability of Compton scattering per unit volume is approximately proportional to the density of the material. Compared to other elements, the absence of neutrons in the hydrogen atom results in an approximate doubling of electron density. Thus, hydrogenous materials have a higher probability of Compton scattering than a nonhydrogenous material of equal mass.

The Photoelectric Effect

In the photoelectric effect, all of the incident photon energy is transferred to an electron, which is ejected from the atom. The kinetic energy of the ejected *photoelectron* (E_e) is equal to the incident photon energy (E_o) minus the binding energy of the orbital electron (E_b) (Fig. 3-9 Left).

$$E_e = E_o - E_b \quad [3-4]$$

In order for photoelectric absorption to occur, the incident photon energy must be greater than or equal to the binding energy of the electron that is ejected. The ejected electron is most likely one whose binding energy is closest to, but less than, the incident photon energy. For example, for photons whose energies exceed the K -shell binding energy, photoelectric interactions with K -shell electrons are most probable. Following a photoelectric interaction, the atom is ionized, with an inner shell electron vacancy. This vacancy will be filled by an electron from a shell with a lower binding energy. This creates another vacancy, which, in turn, is filled by an electron from an even lower binding energy shell. Thus, an electron cascade from outer to inner shells occurs. The difference in binding energy is released as either characteristic x-rays or auger electrons (see Chapter 2). The probability of characteristic x-ray emission decreases as the atomic number of the absorber decreases and thus does not occur frequently for diagnostic energy photon interactions in soft tissue.

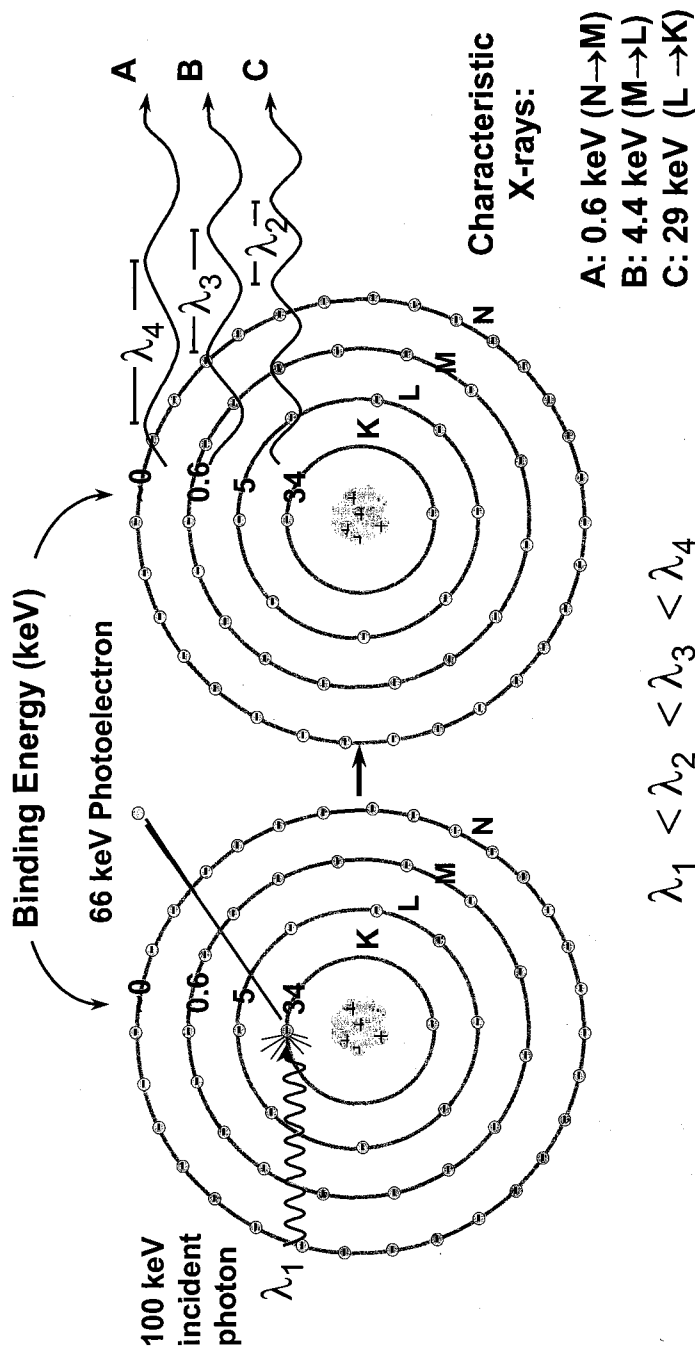


FIGURE 3-9. Photoelectric absorption. **Left:** Diagram shows a 100-keV photon is undergoing photoelectric absorption with an iodine atom. In this case, the K-shell electron is ejected with a kinetic energy equal to the difference between the incident photon energy and the K-shell binding energy of 34 or 66 keV. **Right:** The vacancy created in the K shell results in the transition of an electron from the L shell to the K shell. The difference in their binding energies, (i.e., 34 and 5 keV), results in a 29-keV K_{α} characteristic x-ray. This electron cascade will continue resulting in the production of other characteristic x-rays of lower energies. Note that the sum of the characteristic x-ray energies equals the binding energy of the ejected photoelectrons. Although not shown on this diagram, Auger electrons of various energies could be emitted in lieu of the characteristic x-ray emissions.

Example: The *K*- and *L*-shell electron binding energies of iodine are 34 and 5 keV, respectively. If a 100-keV photon is absorbed by a *K*-shell electron in a photoelectric interaction, the photoelectron is ejected with a kinetic energy equal to $E_p - E_k = 100 - 34 = 66$ keV. A characteristic x-ray or Auger electron is emitted as an outer-shell electron fills the *K*-shell vacancy (e.g., *L* to *K* transition is $34 - 5 = 29$ keV). The remaining energy is released by subsequent cascading events in the outer shells of the atom (i.e., *M* to *L* and *N* to *M* transitions). Note that the total of all the characteristic x-ray emissions in this example equals the binding energy of the *K*-shell photoelectron (Fig. 3-9 right). Thus, photoelectric absorption results in the production of

1. a photoelectron,
2. a positive ion (ionized atom), and
3. characteristic x-rays or Auger electrons.

The probability of photoelectric absorption per unit mass is approximately proportional to Z^3/E^3 , where Z is the atomic number and E is the energy of the incident photon. For example, the photoelectric interaction probability in iodine ($Z = 53$) is $(53/20)^3$ or 18.6 times greater than in calcium ($Z = 20$) for photon of a particular energy.

The benefit of photoelectric absorption in x-ray transmission imaging is that there are no additional nonprimary photons to degrade the image. The fact that the probability of photoelectric interaction is proportional to $1/E^3$ explains, in part, why image contrast decreases when higher x-ray energies are used in the imaging process (see Chapter 10). If the photon energies are doubled, the probability of photoelectric interaction is decreased eightfold: $(1/2)^3 = 1/8$.

Although the probability of the photoelectric effect decreases, in general, with increasing photon energy, there is an exception. For every element, a graph of the probability of the photoelectric effect, as a function of photon energy, exhibits sharp discontinuities called absorption edges (see Fig. 3-11). The probability of interaction for photons of energy just above an absorption edge is much greater than that of photons of energy slightly below the edge. For example, a 33.2-keV x-ray photon is about six times as likely to have a photoelectric interaction with an iodine atom as a 33.1-keV photon.

As mentioned above, a photon cannot undergo a photoelectric interaction with an electron in a particular atomic shell or subshell if the photon's energy is less than the binding energy of that shell or subshell. This causes the dramatic decrease in the probability of photoelectric absorption for photons whose energies are just below the binding energy of a shell. Thus, the photon energy corresponding to an absorption edge is the binding energy of the electrons in that particular shell or subshell. An absorption edge is designated by a letter, representing the atomic shell of the electrons, followed by a number denoting the subshell (e.g., *K*, *L*1, *L*2, *L*3, etc.). The photon energy corresponding to a particular absorption edge increases with the atomic number (Z) of the element. For example, the primary elements comprising soft tissue (H, C, N, and O) have absorption edges below 1 keV. The elements iodine ($Z = 53$) and barium ($Z = 56$), commonly used in radiographic contrast agents to provide enhanced x-ray attenuation, have *K*-absorption edges of 33.2 and 37.4 keV, respectively (Fig. 3-10). The *K*-edge energy of lead ($Z = 82$) is 88 keV.

At photon energies below 50 keV, the photoelectric effect plays an important role in imaging soft tissue. The photoelectric absorption process can be used to

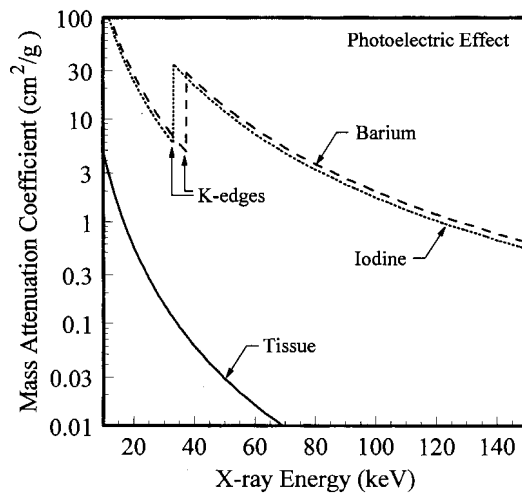


FIGURE 3-10. Photoelectric mass attenuation coefficients for tissue ($Z = 7$), iodine ($Z = 53$), and barium ($Z = 56$) as a function of energy. Abrupt increase in the attenuation coefficients called "absorption edges" occur due to increased probability of photoelectric absorption when the photon energy just exceeds the binding energy of inner shell electrons (e.g., K , L , M ...) thus increasing the number of electrons available for interaction. This process is very significant for high- Z material, such as iodine and barium, in the diagnostic energy range.

amplify differences in attenuation between tissues with slightly different atomic numbers, thereby improving image contrast. This differential absorption is exploited to improve image contrast in the selection of x-ray tube target material and filters in mammography (Chapter 8) and in the use of phosphors containing rare earth elements (e.g., lanthanum and gadolinium) in intensifying screens (Chapter 6). The photoelectric process predominates when lower energy photons interact with high Z materials (Fig. 3-11). In fact, photoelectric absorption is the primary mode of interaction of diagnostic x-rays with screen phosphors, radiographic contrast materials, and bone. Conversely, Compton scattering will predominate at most diagnostic photon energies in materials of lower atomic number such as tissue and air.

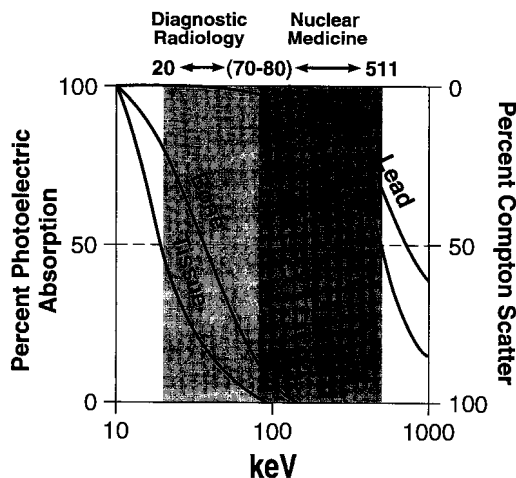


FIGURE 3-11. Graph of the percentage of contribution of photoelectric (left scale) and Compton (right scale) attenuation processes for various materials as a function of energy. When diagnostic energy photons (i.e., diagnostic x-ray effective energy of 20 to 80 keV; nuclear medicine imaging photons of 70 to 511 keV), interact with materials of low atomic number (e.g., soft tissue), the Compton process dominates.

Pair Production

Pair production can only occur when the energies of x- and gamma rays exceed 1.02 MeV. In pair production, an x- or gamma ray interacts with the electric field of the nucleus of an atom. The photon's energy is transformed into an electron-positron pair (Fig. 3-12A). The rest mass energy equivalent of each electron is 0.511 MeV and this is why the energy threshold for this reaction is 1.02 MeV. Photon energy in excess of this threshold is imparted to the electrons as kinetic energy. The electron and positron lose their kinetic energy via excitation and ionization. As discussed previously, when the positron comes to rest, it interacts with a negatively charged electron, resulting in the formation of two oppositely directed 0.511 MeV annihilation photons (Fig. 3-12B).

Pair production is of no consequence in diagnostic x-ray imaging because of the extremely high energies required for it to occur. In fact, pair production does not become significant unless the photon energies greatly exceed the 1.02 MeV energy threshold.

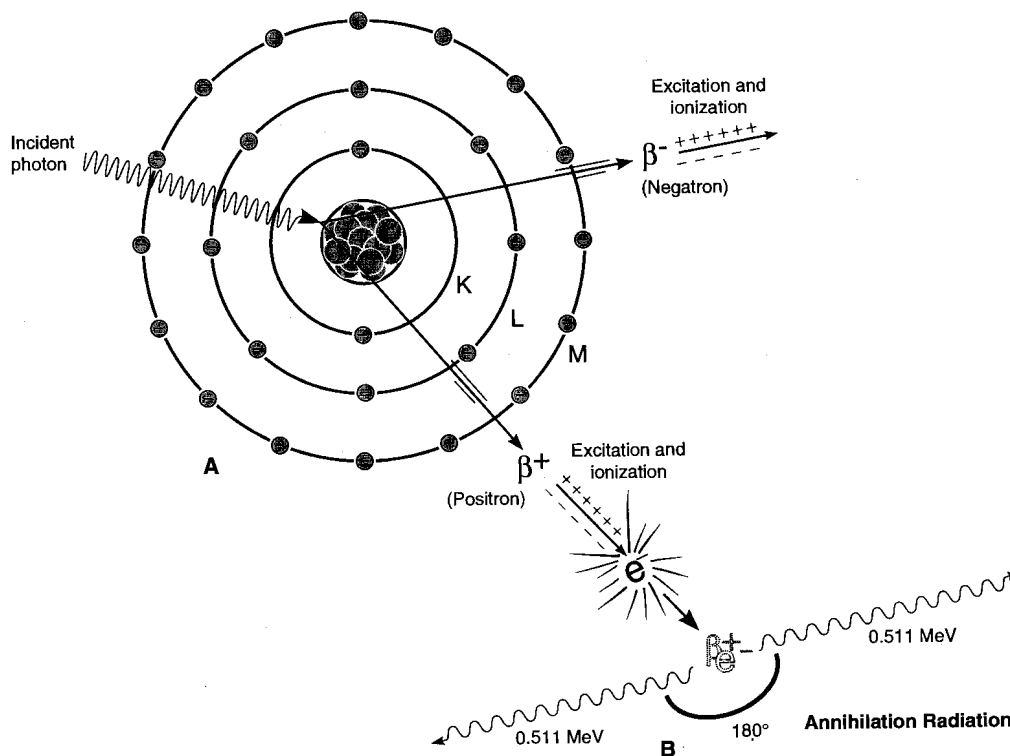


FIGURE 3-12. Pair production. **A:** Diagram illustrates the pair production process in which a high-energy incident photon, under the influence of the atomic nucleus, is converted to the matter and antimatter pair. Both electrons (positron and negatron) expend their kinetic energy by excitation and ionization in the matter they traverse. **B:** However, when the positron comes to rest, it combines with an electron producing the two 511-keV annihilation radiation photons. K, L, and M are electron shells.

3.3 ATTENUATION OF X- AND GAMMA RAYS

Attenuation is the removal of photons from a beam of x- or gamma rays as it passes through matter. Attenuation is caused by both absorption and scattering of the primary photons. The interaction mechanisms discussed in the last section contribute in varying degrees to the attenuation. At low photon energies (<26 keV), the photoelectric effect dominates the attenuation processes in soft tissue. However, as previously discussed, photoelectric absorption is highly dependent on photon energy and the atomic number of the absorber. When higher energy photons interact with low Z materials (e.g., soft tissue), Compton scattering dominates (Fig. 3-13). Rayleigh scattering occurs in medical imaging with low probability, comprising about 10% of the interactions in mammography and 5% in chest radiography. Only at very high photon energies (>1.02 MeV), well beyond the range of diagnostic and nuclear radiology, does pair production contribute to attenuation.

Linear Attenuation Coefficient

The fraction of photons removed from a monoenergetic beam of x- or gamma rays per unit thickness of material is called the *linear attenuation coefficient* (μ), typically expressed in units of inverse centimeters (cm^{-1}). The number of photons removed from the beam traversing a very small thickness Δx can be expressed as:

$$n = \mu N \Delta x \quad [3-5]$$

where n = the number of photons removed from the beam, and N = the number of photons incident on the material.

For example, for 100-keV photons traversing soft tissue, the linear attenuation coefficient is 0.016 mm^{-1} . This signifies that for every 1,000 monoenergetic photons incident upon a 1-mm thickness of tissue, approximately 16 will be removed from the beam, either by absorption or scattering.

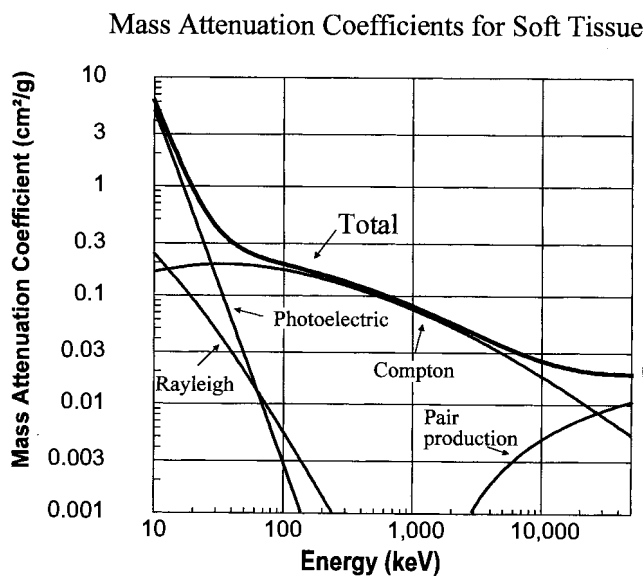


FIGURE 3-13. Graph of the Rayleigh, photoelectric, Compton, pair production, and total mass attenuation coefficients for soft tissue ($Z \approx 7$) as a function of energy.

As the thickness increases, however, the relationship is not linear. For example it would not be correct to conclude from Equation 3-5 that 6 cm of tissue would attenuate 960 (96%) of the incident photons. The attenuation process is continuous from the front surface of the attenuating material to the back exiting surface. To accurately calculate the number of photons removed from the beam using Equation 3-5, multiple calculations utilizing very small thicknesses of material (Δx) would be required. Alternatively, calculus can be employed to simplify this otherwise tedious process.

For a monoenergetic beam of photons incident upon either thick or thin slabs of material, an exponential relationship exists between the number of incident photons (N_0) and those that are transmitted (N) through a thickness x without interaction:

$$N = N_0 e^{-\mu x} \quad [3-6]$$

Thus, using the example above, the fraction of 100-keV photons transmitted through 6 cm of tissue is

$$N/N_0 = e^{-(0.16 \text{ cm}^{-1})(6 \text{ cm})} = 0.38$$

This result indicates that, on average, 380 of the 1,000 incident photons (i.e., 38%) would be transmitted through the tissue without interacting. Thus the actual attenuation ($1 - 0.38$ or 62%) is much lower than would have been predicted from Equation 3-5.

The linear attenuation coefficient is the sum of the individual linear attenuation coefficients for each type of interaction:

$$\mu = \mu_{\text{Rayleigh}} + \mu_{\text{photoelectric effect}} + \mu_{\text{Compton scatter}} + \mu_{\text{pair production}} \quad [3-7]$$

In the diagnostic energy range, the linear attenuation coefficient decreases with increasing energy except at absorption edges (e.g., K -edge). The linear attenuation coefficient for soft tissue ranges from ~ 0.35 to $\sim 0.16 \text{ cm}^{-1}$ for photon energies ranging from 30 to 100 keV.

For a given thickness of material, the probability of interaction depends on the number of atoms the x- or gamma rays encounter per unit distance. The density (ρ , in g/cm^3) of the material affects this number. For example, if the density is doubled, the photons will encounter twice as many atoms per unit distance through the material. Thus, the linear attenuation coefficient is proportional to the density of the material, for instance:

TABLE 3-1. MATERIAL DENSITY, ELECTRONS PER MASS, ELECTRON DENSITY, AND THE LINEAR ATTENUATION COEFFICIENT (AT 50 keV) FOR SEVERAL MATERIALS

Material	Density (g/cm^3)	Electrons per Mass ($\text{e/g}) \times 10^{23}$	Electron Density ($\text{e/cm}^3) \times 10^{23}$	μ @ 50 keV (cm^{-1})
Hydrogen	0.000084	5.97	0.0005	0.000028
Water vapor	0.000598	3.34	0.002	0.000128
Air	0.00129	3.006	0.0038	0.000290
Fat	0.91	3.34	3.04	0.193
Ice	0.917	3.34	3.06	0.196
Water	1	3.34	3.34	0.214
Compact bone	1.85	3.192	5.91	0.573

$$\mu_{\text{water}} > \mu_{\text{ice}} > \mu_{\text{water vapor}}$$

The relationship between material density, electron density, electrons per mass, and the linear attenuation coefficient (at 50 keV) for several materials is shown in Table 3-1.

Mass Attenuation Coefficient

For a given thickness, the probability of interaction is dependent on the number of atoms per volume. This dependency can be overcome by normalizing the linear attenuation coefficient for the density of the material. The linear attenuation coefficient, normalized to unit density, is called the *mass attenuation coefficient*:

$$\begin{aligned} \text{Mass Attenuation Coefficient } (\mu/\rho)[\text{cm}^2/\text{g}] = \\ \frac{\text{Linear Attenuation Coefficient } (\mu)[\text{cm}^{-1}]}{\text{Density of Material } (\rho) [\text{g}/\text{cm}^3]} \end{aligned} \quad [3-8]$$

The linear attenuation coefficient is usually expressed in units of cm^{-1} , whereas the units of the mass attenuation coefficient are cm^2/g .

The mass attenuation coefficient is *independent* of density. Therefore, for a given photon energy:

$$\mu_{\text{water}}/\rho_{\text{water}} = \mu_{\text{ice}}/\rho_{\text{ice}} = \mu_{\text{water vapor}}/\rho_{\text{water vapor}}$$

However, in radiology, we do not usually compare equal masses. Instead, we usually compare regions of an image that correspond to irradiation of adjacent volumes of tissue. Therefore, density, the mass contained within a given volume, plays an important role. Thus, one can radiographically visualize ice in a cup of water due to the density difference between the ice and the surrounding water (Fig. 3-14).

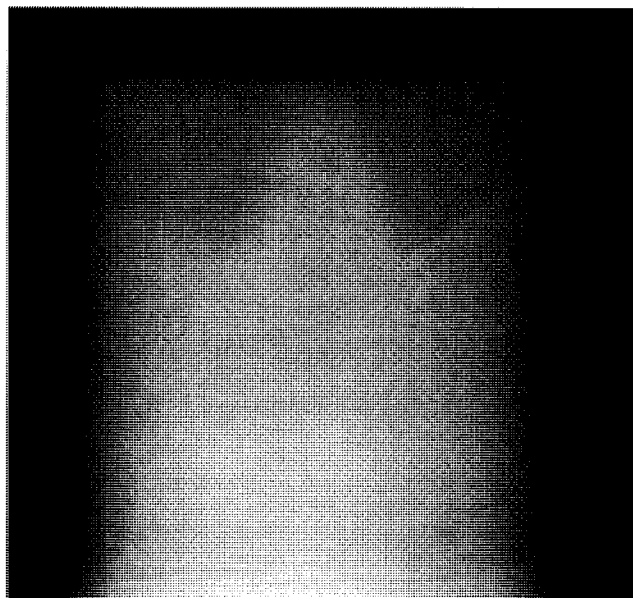


FIGURE 3-14. Radiograph (acquired at 125 kVp with an antiscatter grid) of two ice cubes in a plastic container of water. The ice cubes can be visualized because of their lower electron density relative to that of liquid water. The small radiolucent objects seen at several locations are the result of air bubbles in the water. (From Bushberg JT. The AAPM/RSNA physics tutorial for residents. X-ray interactions. *RadioGraphics* 1998;18:457–468, with permission.)

To calculate the linear attenuation coefficient for a density other than 1 g/cm^3 , then the density of interest ρ is multiplied by the mass attenuation coefficient to yield the linear attenuation coefficient. For example, the mass attenuation coefficient of air, for 60-keV photons, is $0.186 \text{ cm}^2/\text{g}$. At typical room conditions, the density of air is 0.001293 g/cm^3 . Therefore, the linear attenuation coefficient of air at these conditions is:

$$\mu = (\mu/\rho)\rho = (0.186 \text{ cm}^2/\text{g}) (0.001293 \text{ g/cm}^3) = 0.000241 \text{ cm}^{-1}$$

To use the mass attenuation coefficient to compute attenuation, Equation 3-6 can be rewritten as

$$N = N_0 e^{-(\mu/\rho)\rho x} \quad [3-9]$$

Because the use of the mass attenuation coefficient is so common, scientists in this field tend to think of thickness not as a linear distance x (in cm), but rather in terms of mass per unit area ρx (in g/cm^2). The product ρx is called the *mass thickness* or *areal thickness*.

Half Value Layer (Penetrability of Photons)

The half value layer (HVL) is defined as the thickness of material required to reduce the intensity of an x- or gamma-ray beam to one-half of its initial value. The HVL of a beam is an indirect measure of the photon energies (also referred to as the *quality*) of a beam, when measured under conditions of “good” or *narrow-beam geometry*. Narrow-beam geometry refers to an experimental configuration that is designed to exclude scattered photons from being measured by the detector (Fig. 3-15A). In *broad-beam geometry*, the beam is sufficiently wide that a substantial fraction of scattered photons remain in the beam. These scattered photons reaching the detector (Fig. 3-15B) result in an underestimation of the attenuation (i.e., an overestimated

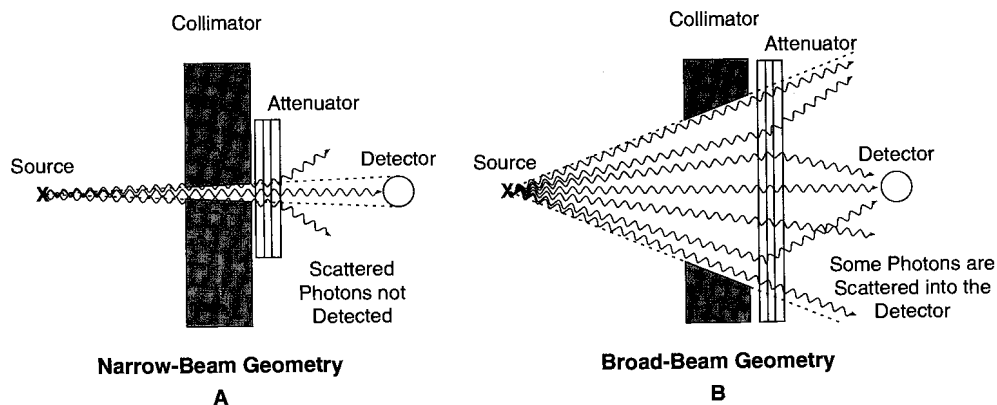


FIGURE 3-15. A: Narrow-beam geometry means that the relationship between the source shield and detector are such that only nonattenuated photons interact with the detector. **B:** In broad-beam geometry attenuated photons may be scattered into the detector; thus the apparent attenuation is lower compared with narrow-beam conditions.

HVL). Most practical applications of attenuation (e.g., patient imaging) occur under broad-beam conditions. The tenth value layer (TVL) is analogous to the HVL, except that it is the thickness of material necessary to reduce the intensity of the beam to a tenth of its initial value. The TVL is often used in x-ray room shielding design calculations (see Chapter 23).

For monoenergetic photons under narrow-beam geometry conditions, the probability of attenuation remains the same for each additional HVL thickness placed in the beam. Reduction in beam intensity can be expressed as $(1/2)^n$ where n equals the number of half value layers. For example, the fraction of monoenergetic photons transmitted through 5 HVLs of material is

$$1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = (1/2)^5 = 1/32 = 0.031 \text{ or } 3.1\%$$

Therefore, 97% of the photons are attenuated (removed from the beam). The HVL of a diagnostic x-ray beam, measured in millimeters of aluminum under narrow-beam conditions, is a surrogate measure of the average energy of the photons in the beam.

It is important to understand the relationship between μ and HVL. In Equation 3-6, N is equal to $N_0/2$ when the thickness of the absorber is 1 HVL. Thus, for a monoenergetic beam:

$$\begin{aligned} N_0/2 &= N_0 e^{-\mu(\text{HVL})} \\ 1/2 &= e^{-\mu(\text{HVL})} \\ \ln(1/2) &= \ln e^{-\mu(\text{HVL})} \\ -0.693 &= -\mu(\text{HVL}) \\ \text{HVL} &= 0.693/\mu \end{aligned} \quad [3-10]$$

For a monoenergetic incident photon beam, the HVL can be easily calculated from the linear attenuation coefficient, and vice versa. For example: given

- a. $\mu = 0.35 \text{ cm}^{-1}$
 $\text{HVL} = 0.693/0.35 \text{ cm}^{-1} = 1.98 \text{ cm}$
- b. $\text{HVL} = 2.5 \text{ mm} = 0.25 \text{ cm}$
 $\mu = 0.693/0.25 \text{ cm} = 2.8 \text{ cm}^{-1}$

The HVL and μ can also be calculated if the percent transmission is measured under narrow-beam geometry.

Example: If a 2-mm thickness of material transmits 25% of a monoenergetic beam of photons, calculate the HVL of the beam.

Step 1. $0.25 = e^{-\mu(0.2 \text{ cm})}$

Step 2. $\ln 0.25 = -\mu(0.2 \text{ cm})$

Step 3. $\mu = (-\ln 0.25)/(0.2 \text{ cm}) = 6.93 \text{ cm}^{-1}$

Step 4. $\text{HVL} = 0.693/\mu = 0.693/6.93 \text{ cm}^{-1} = 0.1 \text{ cm}$

Half value layers for photons from two commonly used diagnostic radiopharmaceuticals [thallium-201 (Tl-201) and technetium-99m (Tc-99m)] are listed for tissue, aluminum, and lead in Table 3-2. Thus, the HVL is a function of (a) photon energy, (b) geometry, and (c) attenuating material.

TABLE 3-2. HALF VALUE LAYERS OF TISSUE, ALUMINUM, AND LEAD FOR X-RAYS AND GAMMA RAYS COMMONLY USED IN DIAGNOSTIC IMAGING

Photon Source	Half Value Layer (mm)		
	Tissue	Aluminum	Lead
70 keV x-rays (Tl 201)	37	11	0.2
140 keV γ -rays (Tc 99m)	44	18	0.3

Tc, technetium; Tl, thallium.

Effective Energy

X-ray beams in radiology are typically composed of a spectrum of energies, dubbed a *polyenergetic* beam. The determination of the HVL in diagnostic radiology is a way of characterizing the hardness of the x-ray beam. The HVL, usually measured in millimeters of aluminum (mm Al) in diagnostic radiology, can be converted to a quantity called the *effective energy*. The effective energy of a polyenergetic x-ray beam is essentially an estimate of the penetration power of the x-ray beam, as if it were a monoenergetic beam. The relationship between HVL (in mm Al) and effective energy is given in Table 3-3.

Mean Free Path

One cannot predict the range of a single photon in matter. In fact, the range can vary from zero to infinity. However, the average distance traveled before interaction can be calculated from the linear attenuation coefficient or the HVL of the beam. This length, called the *mean free path* (MFP) of the photon beam, is

TABLE 3-3. HALF VALUE LAYER (HVL) AS A FUNCTION OF THE EFFECTIVE ENERGY OF AN X-RAY BEAM

HVL (mm Al)	Effective Energy (keV)
0.26	14
0.39	16
0.55	18
0.75	20
0.98	22
1.25	24
1.54	26
1.90	28
2.27	30
3.34	35
4.52	40
5.76	45
6.97	50
9.24	60
11.15	70
12.73	80
14.01	90
15.06	100

Al, aluminum.

$$\text{MFP} = \frac{1}{\mu} = \frac{1}{0.693/\text{HVL}} = 1.44 \text{ HVL} \quad [3-11]$$

Beam Hardening

The lower energy photons of the polyenergetic x-ray beam will preferentially be removed from the beam while passing through matter. The shift of the x-ray spectrum to higher effective energies as the beam transverses matter is called *beam hardening* (Fig. 3-16). Low-energy (soft) x-rays will not penetrate most tissues in the body; thus their removal reduces patient exposure without affecting the diagnostic quality of the exam. X-ray machines remove most of this soft radiation with filters, thin plates of aluminum, or other materials placed in the beam. This filtering will result in an x-ray beam with a higher effective energy, and thus a larger HVL.

The homogeneity coefficient is the ratio of the first to the second HVL and describes the polyenergetic character of the beam. The first HVL is the thickness that reduces the incident intensity to 50% and the second HVL reduces it to 25% of its original intensity [i.e., $(0.5)(0.5) = 0.25$]. A monoenergetic source of gamma rays has a homogeneity coefficient equal to 1.

The maximal x-ray energy of a polyenergetic spectrum can be estimated by monitoring the homogeneity coefficient of two heavily filtered beams (e.g., 15th and 16th HVLs). As the coefficient approaches 1, the beam is essentially monoenergetic. Measuring μ for the material in question under heavy filtration conditions and matching to known values of μ for monoenergetic beams gives an approximation of the maximal energy.

The effective (average) energy of an x-ray beam from a typical diagnostic x-ray tube is one-third to one-half the maximal value and gives rise to an “effective μ ” —

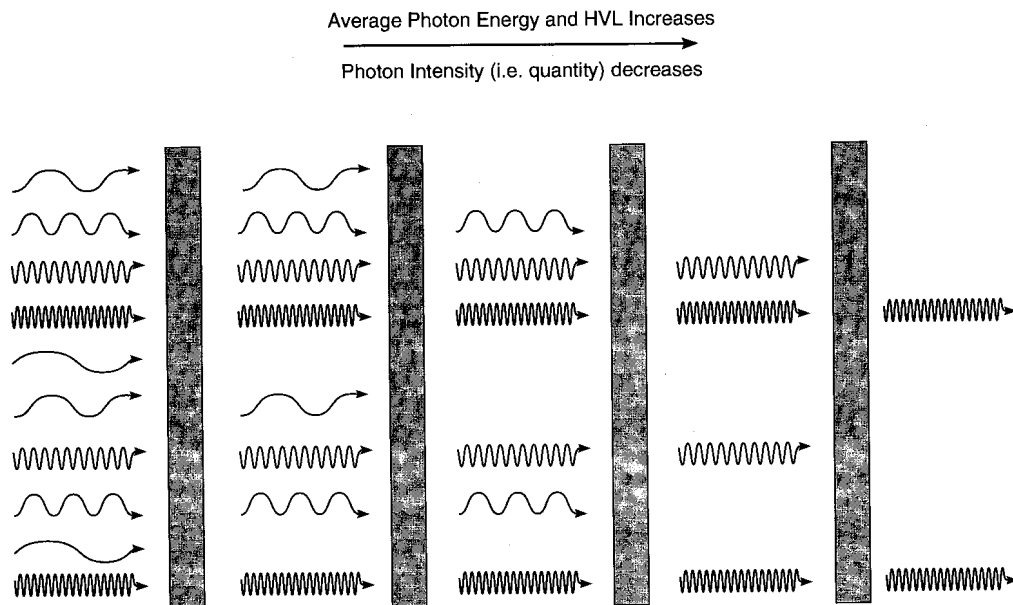


FIGURE 3-16. Beam hardening results from preferential absorption of lower energy photons as they traverse matter.

the attenuation coefficient that would be measured if the x-ray beam were monoenergetic at that “effective” energy.

3.4 ABSORPTION OF ENERGY FROM X- AND GAMMA RAYS

Fluence, Flux, and Energy Fluence

The number of photons (or particles) passing through a unit cross-sectional area is referred to as the *fluence* and is typically expressed in units of cm^{-2} . The fluence is given the symbol Φ :

$$\Phi = \frac{\text{Photons}}{\text{Area}} \quad [3-12]$$

The fluence rate (e.g., the rate at which photons or particles pass through a unit area per unit time) is called the *flux*. The flux, given the symbol $\dot{\Phi}$, is simply the fluence per unit time.

$$\dot{\Phi} = \frac{\text{Photons}}{\text{Area Time}} \quad [3-13]$$

The flux is useful in areas where the photon beam is on for extended periods of time, such as in fluoroscopy. The flux has the units of $\text{cm}^{-2} \text{sec}^{-1}$.

The amount of energy passing through a unit cross-sectional area is referred to as the *energy fluence*. For a monoenergetic beam of photons, the energy fluence (Ψ) is simply the product of the fluence (Φ) and the energy per photon (E).

$$\Psi = \left(\frac{\text{Photons}}{\text{Area}} \right) \times \left(\frac{\text{Energy}}{\text{Photon}} \right) \Phi E \quad [3-14]$$

The units of Ψ are energy per unit area, keV per cm^2 , or joules per m^2 . For a polyenergetic spectrum, the total energy in the beam is tabulated by multiplying the number of photons at each energy by that energy. The energy fluence rate or energy flux is the energy fluence per unit time.

Kerma

As a beam of indirectly ionizing radiation (e.g., x- or gamma rays or neutrons) passes through a medium, it deposits energy in the medium in a two-stage process:

Step 1. Energy carried by the photons (or other indirectly ionizing radiation) is transformed into kinetic energy of charged particles (such as electrons). In the case of x- and gamma rays, the energy is transferred by the photoelectric effect, Compton effect, and, for very high energy photons, pair production.

Step 2. The directly ionizing charged particles deposit their energy in the medium by excitation and ionization. In some cases, the range of the charged particles is sufficiently large that energy deposition is some distance away from the initial interactions.

Kerma (K) is an acronym for *k*inetic energy released in *m*atter. Kerma is defined as the kinetic energy transferred to charged particles by indirectly ionizing radiation, per mass matter, as described in Step 1 above. Kerma is expressed in units of J/kg or gray (defined below).

For x- and gamma rays, kerma can be calculated from the mass energy transfer coefficient of the material and the energy fluence.

Mass Energy Transfer Coefficient

The mass energy transfer coefficient is given the symbol:

$$\left(\frac{\mu_{tr}}{\rho_o} \right)$$

The mass energy transfer coefficient is the mass attenuation coefficient multiplied by the fraction of the energy of the interacting photons that is transferred to charged particles as kinetic energy. As was mentioned above, energy deposition in matter by photons is largely delivered by the production of energetic charged particles. The energy in scattered photons that escape the interaction site is not transferred to charged particles in the volume of interest. Furthermore, when pair production occurs, 1.02 MeV of the incident photon's energy is required to produce the electron-positron pair and only the remaining energy ($E_{\text{photon}} - 1.02 \text{ MeV}$) is given to the electron and positron as kinetic energy. Therefore, the mass energy transfer coefficient will always be less than the mass attenuation coefficient.

For 20-keV photons in tissue, for example, the ratio of the energy transfer coefficient to the attenuation coefficient (μ_{tr}/μ) is 0.68, but this reduces to 0.18 for 50-keV photons, as the amount of Compton scattering increases relative to photoelectric absorption.

Calculation of Kerma

For a monoenergetic photon beams with an energy fluence Ψ and energy E , the kerma K is given by:

$$K = \Psi \left(\frac{\mu_{tr}}{\rho_o} \right)_E \quad [3-15]$$

where

$$\left(\frac{\mu_{tr}}{\rho_o} \right)_E$$

is the mass energy transfer coefficient of the absorber at energy E . The SI units of energy fluence are J/m^2 , and the SI units of the mass energy transfer coefficient are m^2/kg , and thus their product, kerma, has units of J/kg .

Absorbed Dose

The quantity *absorbed dose* (D) is defined as the energy (ΔE) deposited by ionizing radiation per unit mass of material (Δm):

$$\text{Dose} = \frac{\Delta E}{\Delta m} \quad [3-16]$$

Absorbed dose is defined for all types of ionizing radiation. The SI unit of absorbed dose is the *gray* (Gy). One gray is equal to 1 J/kg. The traditional unit of

absorbed dose is the *rad* (an acronym for radiation absorbed dose). One rad is equal to 0.01 J/kg. Thus, there are 100 rads in a gray and 1 rad = 10 mGy.

If the energy imparted to charged particles is deposited locally and the bremsstrahlung losses are negligible, the absorbed dose will be equal to the kerma.

For x- and gamma rays, the absorbed dose can be calculated from the mass energy absorption coefficient and the energy fluence of the beam.

Mass Energy Absorption Coefficient

The mass energy transfer coefficient discussed above describes the fraction of the mass attenuation coefficient that gives rise to the initial kinetic energy of electrons in a small volume of absorber. These energetic electrons may subsequently produce bremsstrahlung radiation (x-rays), which can escape the small volume of interest. Thus, the mass energy absorption coefficient is slightly smaller than the mass energy transfer coefficient. For the energies used in diagnostic radiology and for low Z absorbers (air, water, tissue), of course, the amount of radiative losses (bremsstrahlung) are very small. Thus, for diagnostic radiology:

$$\left(\frac{\mu_{en}}{\rho_o}\right) \cong \left(\frac{\mu_{tr}}{\rho_o}\right)$$

The mass energy absorption coefficient is useful when energy deposition calculations are to be made.

Calculation of Dose

The distinction between kerma and dose is slight for the relatively low x-ray energies used in diagnostic radiology. The dose in any material given by

$$D = \Psi \left(\frac{\mu_{en}}{\rho_o}\right)_E \quad [3-17]$$

The difference between kerma and dose for air is that kerma is defined using the mass energy transfer coefficient, whereas dose is defined using the mass energy absorption coefficient. The mass energy transfer coefficient includes energy transferred to charged particles, but these energetic charged particles (mostly electrons) in the absorber will radiate bremsstrahlung radiation, which can exit the small volume of interest. The coefficient

$$\left(\frac{\mu_{en}}{\rho_o}\right)_E$$

takes into account the radiative losses, and thus

$$\left(\frac{\mu_{tr}}{\rho_o}\right)_E \geq \left(\frac{\mu_{en}}{\rho_o}\right)_E$$

Exposure

The amount of electrical charge (ΔQ) produced by ionizing electromagnetic radiation per mass (Δm) of air is called *exposure* (X):

$$X = \frac{\Delta Q}{\Delta m} \quad [3-18]$$

Exposure is in the units of charge per mass, i.e., coulombs per kg or C/kg. The historical unit of exposure is the roentgen (abbreviated R), which is defined as

$$1 \text{ R} = 2.58 \times 10^{-4} \text{ C/kg (exactly)}$$

Radiation fields are often expressed as an exposure rate (R/hr or mR/min). The output intensity of an x-ray machine is measured and expressed as an exposure (R) per unit of current times exposure duration (milliamperere second or mAs) under specified operating conditions (e.g., 5 mR/mAs at 70 kVp for a source/image distance of 100 cm, and with an x-ray beam filtration equivalent to 2 mm Al).

Exposure is a useful quantity because ionization can be directly measured with standard air-filled radiation detectors, and the effective atomic numbers of air and soft tissue are approximately the same. Thus exposure is nearly proportional to dose in soft tissue over the range of photon energies commonly used in radiology. However, the quantity of exposure is limited in that it applies only to the interaction of ionizing photons (not charged particle radiation) in air (not any other substance). Furthermore, it is very difficult to measure exposure for photons of energy greater than about 3 MeV.

There is a relationship between the amount of ionization in air and the absorbed dose in rads for a given photon energy and absorber. The absorbed dose in rad per roentgen of exposure is known as the “roentgen-to-rad conversion factor” and is approximately equal to 1 for soft tissue in the diagnostic energy range (Fig. 3-17). The roentgen-to-rad conversion factor approaches 4 for bone at photon ener-

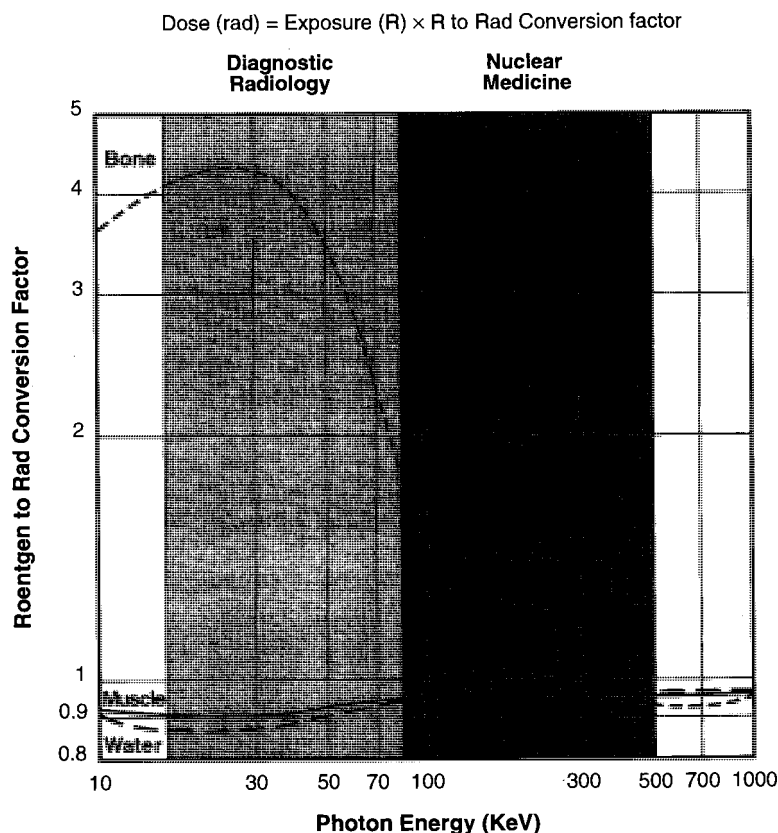


FIGURE 3-17. The roentgen-to-rad conversion factor versus photon energy for water, muscle, and bone.

gies below 100 keV due to an increase in the photoelectric absorption cross section for higher Z materials.

The exposure can be calculated from the dose to air. Let the ratio of the dose (D) to the exposure (X) in air be W , and substituting in the above expressions for D and X yields

$$W = \frac{D}{X} = \frac{\Delta E}{\Delta Q} \quad [3-19]$$

W , the average energy deposited per ion pair in air, is approximately constant as a function of energy. The value of W is 33.97 J/C.

W is the conversion factor between exposure in air and dose in air. In terms of the traditional unit of exposure, the roentgen, the dose to air is:

$$D_{\text{air}} = W \left(\frac{33.97 \text{ J}}{\text{C}} \right) \times X \left(\frac{2.58 \times 10^{-4} \text{ C}}{\text{kg}} \right)$$

or

$$D_{\text{air}} (\text{Gy}) = 0.00876 \left(\frac{\text{Gy}}{\text{R}} \right) \times X (\text{R}) \quad [3-20]$$

Thus, one R of exposure results in 8.76 mGy (0.876 rad) of air dose. The quantity exposure is still in common use in the United States, but the equivalent SI quantities of air dose or air kerma are predominantly used in many other countries. For clarity, these conversions are

$$D_{\text{air}} (\text{mGy}) = 8.76 \times X (\text{R}) \quad [3-21]$$

$$D_{\text{air}} (\mu\text{Gy}) = 8.76 \times X (\text{mR}) \quad [3-22]$$

3.5 IMPARTED ENERGY, EQUIVALENT DOSE, AND EFFECTIVE DOSE

Imparted Energy

The total amount of energy deposited in matter, called the *imparted energy*, is the product of the dose and the mass over which the energy is imparted. The unit of imparted energy is the joule:

$$(\text{J/kg}) \times \text{kg} = \text{J} \quad [3-23]$$

For example, assume each 1-cm slice of a head computed tomography (CT) scan delivers a 30 mGy (3 rad) dose to the tissue in the slice. If the scan covers 15 cm, the dose to the irradiated volume (ignoring scatter from adjacent slices) is still 30 mGy (3 rad); however, the imparted (absorbed) energy is approximately 15 times that of a single scan slice.

Equivalent Dose

Not all types of radiation cause the same biologic damage per unit dose. To modify the dose to reflect the relative effectiveness of the type of radiation in producing biologic damage, a *radiation weighting factor* (w_R) was established by the International Commission on Radiological Protection (ICRP) in publication 60 (1990). High

TABLE 3-4. RADIATION WEIGHTING FACTORS (w_R) FOR VARIOUS TYPES OF RADIATION^a

Type of Radiation	Radiation Weighting Factor (w_R)
X-rays, gamma rays, beta particles, and electrons ^b	1
Protons (>2 MeV)	5
Neutrons (energy dependent)	5–20
Alpha particles and other multiple-charged particles	20

^aFor radiations principally used in medical imaging (x-rays, gamma rays, beta particles) $w_R = 1$; thus the absorbed dose and equivalent dose are equal (i.e., 1 Gy = 1 Sv). Adapted from *1990 Recommendations of the International Commission on Radiological Protection*. ICRP publication no. 60. Oxford: Pergamon, 1991.

^b $w_R = 1$ for electrons of all energies except for auger electrons emitted from nuclei bound to DNA.

LET radiations that produce dense ionization tracks cause more biologic damage per unit dose than low LET radiations and thus have higher radiation weighting factors. The product of the absorbed dose (D) and the radiation weighing factor is the *equivalent dose* (H).

$$H = D w_R \quad [3-24]$$

The SI unit for equivalent dose is the *sievert* (Sv). Radiations used in diagnostic imaging (x-rays, gamma rays, and electrons) have a w_R of 1: thus 1 mGy = 1 mSv. For heavy charged particles such as alpha particles, the LET is much higher and thus the biologic damage and the associated w_R is much greater (Table 3-4). For example, 10 mGy from alpha radiation may have the same biologic effectiveness as 200 mGy of x-rays.

The quantity H replaced an earlier but similar quantity, the *dose equivalent*, which is the product of the absorbed dose and the *quality factor* (Q) (Equation 3-25). The quality factor is analogous to w_R .

$$H = DQ \quad [3-25]$$

The traditional unit for both the dose equivalent and the equivalent dose is the rem. A sievert is equal to 100 rem, and 1 rem is equal to 10 mSv.

Effective Dose

Not all tissues are equally sensitive to the effects of ionizing radiation. Tissue weighting factors (w_T) were established in ICRP publication 60 to assign a particular organ or tissue (T) the proportion of the detriment from stochastic effects (e.g., cancer and genetic effects, discussed further in Chapter 25) resulting from irradiation of that tissue compared to uniform whole-body irradiation. These tissue weighting factors are shown in Table 3-5. The sum of the products of the equivalent dose to each organ or tissue irradiated (H_T) and the corresponding weighting factor (w_T) for that organ or tissue is called the *effective dose* (E). The effective dose (Equation 3-26) is expressed in the same units as the equivalent dose (sievert or rem).

TABLE 3-5. TISSUE WEIGHTING FACTORS ASSIGNED BY THE INTERNATIONAL COMMISSION ON RADIOLOGICAL PROTECTION

Tissue or Organ	Tissue Weighting Factor, w_T
Gonads	0.20
Bone marrow (red)	0.12
Colon	0.12
Lung	0.12
Stomach	0.12
Bladder	0.05
Breast	0.05
Liver	0.05
Esophagus	0.05
Thyroid	0.05
Skin	0.01 ^a
Bone surface	0.01
Remainder	0.05 ^{b,c}
Total	1.00

^aApplied to the mean equivalent dose over the entire skin.

^bFor purposes of calculation, the remainder is composed of the following additional tissues and organs: adrenals, brain, upper large intestine, small intestine, kidney, muscle, pancreas, spleen, thymus, and uterus.

^cIn those exceptional cases in which a single one of the remainder tissues or organs receives an equivalent dose in excess of the highest dose in any of the 12 organs for which weighting factor is specified, a weighting factor of 0.025 should be applied to that tissue or organ and weighting factor of 0.025 to the average dose in the rest of the remainder as defined above.

Adapted from 1990 *Recommendations of the International Commission on Radiological Protection*. ICRP publication no. 60. Oxford: Pergamon, 1991.

$$E(Sv) = \sum w_T \times H_T(Sv) \quad [3-26]$$

The w_T values were developed for a reference population of equal numbers of both sexes and a wide range of ages. Before 1990, the ICRP had established different w_T values (ICRP publication 26, 1977) that were applied as shown in Equation 3-26, the product of which was referred to as the *effective dose equivalent* (H_E). Many regulatory agencies, including the U.S. Nuclear Regulatory Commission (NRC), have not as yet adopted the ICRP publication 60 w_T values and are currently using the ICRP publication 26 dosimetric quantities in their regulations.

Summary

The roentgen, rad, and rem are still commonly used in the United States; however, most other countries and the majority of the world's scientific literature use SI units. The SI units and their traditional equivalents are summarized in Table 3-6.

TABLE 3-6. RADIOLOGICAL QUANTITIES, SYSTEM INTERNATIONAL (SI) UNITS, AND TRADITIONAL UNITS

Quantity	Description of Quantity	SI Units (Abbreviations) and Definitions	Traditional Units (Abbreviations) and Definitions	Symbol	Definitions and Conversion Factors
Exposure	Amount of ionization per mass of air due to x- and gamma rays	C kg ⁻¹	Roentgen (R)	X	1 R = 2.58×10^{-4} C kg ⁻¹ 1 R = 8.708 mGy air kerma @ 30 kVp 1 R = 8.767 mGy air kerma @ 60 kVp 1 R = 8.883 mGy air kerma @ 100 kVp
Absorbed dose	Amount of energy imparted by radiation per mass	Gray (Gy) 1 Gy = J kg ⁻¹	rad 1 rad = 0.01 J kg ⁻¹	D	1 rad = 10 mGy 100 rad = 1 Gy
Kerma	Kinetic energy transferred to charged particles per unit mass	Gray (Gy) 1 Gy = J kg ⁻¹	—	K	—
Air kerma	Kinetic energy transferred to charged particles per unit mass of air	Gray (Gy) 1 Gy = J kg ⁻¹	—	K _{air}	1 mGy = 0.115 R @ 30 kVp 1 mGy = 0.114 R @ 60 kVp 1 mGy = 0.113 R @ 100 kVp 1 mGy $\approx$ 0.014 rad (dose to skin) 1 mGy $\approx$ 1.4 mGy (dose to skin) Dose (J kg ⁻¹) $\times$ mass (kg) = J
Imparted energy	Total radiation energy imparted to matter	Joule (J)	—	D _i	—
Equivalent dose (defined by ICRP in 1990 to replace dose equivalent)	A measure of radiation specific biologic damage in humans	Sievert (Sv)	rem	H	$H = w_R D$ 1 rem = 10 mSv 100 rem = 1 Sv
Dose equivalent (defined by ICRP in 1977)	A measure of radiation specific biologic damage in humans	Sievert (Sv)	rem	H	$H = Q D$ 1 rem = 10 mSv 100 rem = 1 Sv
Effective dose (defined by ICRP in 1990 to replace effective dose equivalent)	A measure of radiation and organ system specific damage in humans	Sievert (Sv)	rem	E	$E = \sum_T w_T H_T$
Effective dose equivalent (defined by ICRP in 1977)	A measure of radiation and organ system specific damage in humans	Sievert (Sv)	rem	H _E	$H_E = \sum_T w_T H_T$
Activity	Amount of radioactive material expressed as the nuclear transformation rate.	Becquerel (Bq) (sec ⁻¹)	Curie (Ci)	A	1 Ci = 3.7×10^{10} Bq 37 kBq = 1 μ Ci 37 MBq = 1 mCi 37 GBq = 1 Ci

ICRP, International Commission on Radiological Protection.

SUGGESTED READING

- Bushberg JT. The AAPM/RSNA physics tutorial for residents. X-ray interactions. *RadioGraphics* 1998;18:457–468.
- International Commission on Radiation Units and Measurements. *Fundamental quantities and units for ionizing radiation*. ICRU report 60. Bethesda, MD: ICRU, 1998.
- International Commission on Radiological Protection. *1990 recommendations of the International Commission on Radiological Protection*. Annals of the ICRP 21 no. 1-3, publication 60. Philadelphia: Elsevier Science, 1991.
- Johns HE, Cunningham JR. *The physics of radiology*, 4th ed. Springfield, IL: Charles C Thomas, 1983.
- McKetty MH. The AAPM/RSNA physics tutorial for residents. X-ray attenuation. *RadioGraphics* 1998;18:151–163.

COMPUTERS IN MEDICAL IMAGING

Computers were originally designed to perform mathematical computations and other data processing tasks very quickly. Since then, they have come to be used for many other purposes, including information display, information storage, and, in conjunction with computer networks, information transfer and communications.

Computers were introduced in medical imaging in the early 1970s and have become increasingly important since that time. Their first uses in medical imaging were mostly in nuclear medicine, where they were used to acquire series of images depicting the organ-specific kinetics of radiopharmaceuticals. Today, computers are essential to many imaging modalities, including x-ray computed tomography (CT), magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT), positron emission tomography (PET), and digital radiography.

Any function that can be performed by a computer can also be performed by a hard-wired electronic circuit. The advantage of the computer over a hard-wired circuit is its flexibility. The function of the computer can be modified merely by changing the program that controls the computer, whereas modifying the function of a hard-wired circuit usually requires replacing the circuit. Although the computer is a very complicated and powerful data processing device, the actual operations performed by a computer are very simple. The power of the computer is mainly due to its speed in performing these operations and its ability to store large volumes of data.

In the next section of this chapter, the storage and transfer of data in digital form are discussed. The following sections describe the components of a computer; factors that determine the performance of a computer; computer software; and the acquisition, storage, and display of digital images. Computer networks, picture archiving, and communications systems (PACS), and teleradiology are discussed in detail in Chapter 17.

4.1 STORAGE AND TRANSFER OF DATA IN COMPUTERS

Number Systems

Our culture uses a number system based on ten, probably because humans have five fingers on each hand and number systems having evolved from the simple act of counting on the fingers. Computers use the binary system for the electronic storage and manipulation of numbers.

Decimal Form (Base 10)

In the decimal form, the ten digits 0 through 9 are used to represent numbers. To represent numbers greater than 9, several digits are placed in a row. The value of each digit in a number depends on its position in the row; the value of a digit in any position is *ten* times the value it would represent if it were shifted one place to the right. For example, the decimal number 3,506 actually represents:

$$(3 \times 10^3) + (5 \times 10^2) + (0 \times 10^1) + (6 \times 10^0)$$

where $10^1 = 10$ and $10^0 = 1$. The leftmost digit in a number is called the *most significant digit* and the rightmost digit is called the *least significant digit*.

Binary Form (Base 2)

In binary form, the two digits 0 and 1 are used to represent numbers. Each of these two digits, by itself, has the same meaning that it has in the decimal form. To represent numbers greater than 1, several digits are placed in a row. The value of each digit in a number depends on its position in the row; the value of a digit in any position is *two* times the value it would represent if it were shifted one place to the right. For example, the binary number 1101 represents:

$$(1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0)$$

where $2^3 = 8$, $2^2 = 4$, $2^1 = 2$, and $2^0 = 1$. To count in binary form, 1 is added to the least significant digit of a number. If the least significant digit is 1, it is replaced by 0 and 1 is added to the next more significant digit. If several contiguous digits on the right are 1, each is replaced by 0 and 1 is added to the least significant digit that was not 1. Counting the binary system is illustrated in Table 4-1.

Conversions Between Decimal and Binary Forms

To convert a number from binary to decimal form, the binary number is expressed as a series of powers of two and the terms in the series are added. For example, to convert the binary number 101010 to decimal form:

$$101010 \text{ (binary)} = (1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (0 \times 2^0)$$

TABLE 4-1. NUMBERS IN DECIMAL AND BINARY FORMS

Decimal	Binary	Decimal	Binary
0	0	8	1000
1	1	9	1001
2	10	10	1010
3	11	11	1011
4	100	12	1100
5	101	13	1101
6	110	14	1110
7	111	15	1111
		16	10000

TABLE 4-2. CONVERSION OF 42 (DECIMAL) INTO BINARY FORM^a

Division	Result	Remainder	
42/2 =	21	0	Least significant digit
21/2 =	10	1	
10/2 =	5	0	
5/2 =	2	1	
2/2 =	1	0	Most significant digit
1/2 =	0	1	

^aIt is repeatedly divided by 2, with the remainder recorded after each division. The binary equivalent of 42 (decimal) is therefore 101010.

$$= (1 \times 32) + (1 \times 8) + (1 \times 2) = 42 \text{ (decimal)}.$$

To convert a number from decimal into binary representation, it is repeatedly divided by two. Each division determines one digit of the binary representation, starting with the least significant. If there is no remainder from the division, the digit is 0; if the remainder is 1, the digit is 1. The conversion of 42 (decimal) into binary form is illustrated in Table 4-2.

Considerations Regarding Number Systems

Whenever it is not clear which form is being used, the form will be written in parentheses after the number. If the form is not specified, it can usually be assumed that a number is in decimal form. It is important not to confuse the binary representation of a number with its more familiar decimal representation. For example, 10 (binary) and 10 (decimal) represent different numbers, although they look alike. On the other hand, 1010 (binary) and 10 (decimal) represent the same number. The only numbers that appear the same in both systems are 0 and 1.

Digital Representation of Data

Bits, Bytes, and Words

Computer memory and storage consist of many elements called *bits* (for *binary digits*), each of which can be in one of two states. A bit is similar to a light switch, because a light switch can only be on or off. In most computer memory, each bit is a transistor; *on* is one state and *off* is the other. In magnetic storage media, such as magnetic disks and magnetic tape, a bit is a small portion of the disk or tape that may be magnetized.

Bits are grouped into *bytes*, each consisting of eight bits. The capacity of a computer memory or storage device is usually described in kilobytes, megabytes, gigabytes, or terabytes (Table 4-3). Bits are also grouped into *words*. The number of bits in a word depends on the computer system; 16-, 32-, and 64-bit words are common.

Digital Representation of Different Types of Data

General-purpose computers must be able to store and process several types of data in digital form. For example, if a computer is to execute a word processing program

TABLE 4-3. UNITS TO DESCRIBE COMPUTER STORAGE CAPACITY^a

1 kilobyte (kB)	= 2^{10} bytes = 1024 bytes $\approx$ a thousand bytes
1 megabyte (MB)	= 2^{20} bytes = 1024 kilobytes = 1,048,576 bytes $\approx$ a million bytes
1 gigabyte (GB)	= 2^{30} bytes = 1024 megabytes = 1,073,741,824 bytes $\approx$ a billion bytes
1 terabyte (TB)	= 2^{40} bytes = 1024 gigabytes = 1,099,511,627,776 $\approx$ a trillion bytes

^aNote that the prefixes *kilo-*, *mega-*, *giga-*, and *tera-* have slightly different meanings when used to describe computer storage capacity than in standard scientific usage.

or a program that stores and retrieves patient data, it must be able to represent alphanumeric data (text), such as a patient's name, in digital form. Computers must also be able to represent numbers in digital form. Most computers have several data formats for numbers. For example, one computer system provides formats for 1-, 2-, 4-, and 8-byte positive integers; 1-, 2-, 4-, and 8-byte signed integers ("signed" meaning that they may have positive or negative values); and 4- and 8-byte floating-point numbers (to be discussed shortly).

When numbers can assume very large or very small values or must be represented with great accuracy, formats requiring a large amount of storage per number must be used. However, if numbers are restricted to integers within a limited range of values, considerable savings in storage and processing time may be realized. For instance, the gray-scale values in ultrasound images typically range from 0 (black) to 255 (white), a range of 256 numbers. As will be seen in the following section, any gray-scale value within this range can be represented by only eight bits. The following sections describe schemes for the digital representation of different types of data.

Storage of Positive Integers

As discussed above, computer memory and storage consist of many bits. Each bit can be in one of two states and can therefore represent the numbers 0 or 1. Two bits have four possible configurations (00, 01, 10, or 11) that in decimal form are 0, 1, 2, and 3. Three bits have eight possible configurations (000, 001, 010, 011, 100, 101, 110, or 111) that can represent the decimal numbers 0, 1, 2, 3, 4, 5, 6, and 7. In general, N bits have 2^N possible configurations and can represent integers from 0 to $2^N - 1$. One byte can therefore store integers from 0 to 255 and a 16-bit word can store integers from 0 to 65,535 (Table 4-4).

TABLE 4-4. NUMBER OF BITS REQUIRED TO STORE INTEGERS

Number of Bits	Possible Configurations	Number of Configurations	Represent Integers (Decimal Form)
1	0,1	2	0,1
2	00,01,10,11	4	0,1,2,3
3	000,001,010,011,100,101,110,111	8	0,1,2,3,4,5,6,7
8	00000000 to 11111111	256	0 to 255
16	0000000000000000 to 1111111111111111	65,536	0 to 65,535
N		2^N	0 to $2^N - 1$

Binary Representation of Signed Integers

The previous discussion dealt only with positive integers. It is often necessary for computers to manipulate integers that can have positive or negative values. There are many ways to represent signed integers in binary form. The simplest method is to reserve the first bit of a number for the sign of the number. Setting the first bit to 0 can indicate that the number is positive, whereas setting it to 1 indicates that the number is negative. For an 8-bit number, 11111111 (binary) = -127 (decimal) is the most negative number and 01111111 (binary) = 127 (decimal) is the largest positive number. Most computers use a different scheme called “twos complement notation” to represent signed integers, which simplifies the circuitry needed to add positive and negative integers.

Floating-Point Form

Computers used for scientific purposes must be able to manipulate very large numbers, such as Avogadro’s number (6.022×10^{23} molecules per gram-mole) and very small numbers, such as the mass of a proton (1.673×10^{-27} kilograms). These numbers are usually represented in floating-point form. Floating-point form is similar to scientific notation, in which a number is expressed as a decimal fraction times ten raised to a power. A number can be also written as a binary fraction times two to a power. For example, Avogadro’s number can be written as

$$0.11111111 \text{ (binary)} \times 2^{01001111} \text{ (binary)}.$$

When a computer stores the number in floating-point form, it stores the pair of signed binary integers, 11111111 and 01001111.

Binary Representation of Text

It is often necessary for computers to store and manipulate alphanumeric data, such as a patient’s name or the text of this book. One common method for representing alphanumeric data in binary form is the American Standard Code for Information Interchange (ASCII). Each character is stored in one byte. The byte values from 00000000 to 01111111 (binary) are used to define 128 characters, including upper- and lowercase letters, the digits 0 through 9, many punctuation marks, and several carriage-control characters such as line feed. For example, the carriage-control character “carriage return” is represented by 00001101, the uppercase letter “A” is represented by 01000001, the comma is represented by 00111010, and the digit “2” is represented by 00110010. Several other methods for representing text in binary form that permit more than 128 characters to be represented are also in use.

Transfers of Data in Digital Form

Data are transferred between the various components of a computer, such as the memory and central processing unit, in binary form. A voltage of fixed value (such as 5 V) and fixed duration on a wire can be used to represent the binary number 1 and a voltage of 0 V for the same duration can represent 0. A group of such voltage pulses can be used to represent a number, an alphanumeric character, or other unit of information. (This scheme is called “unipolar” digital encoding. Many other digital encoding schemes exist.)

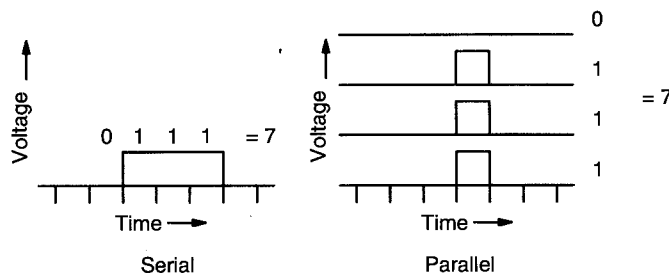


FIGURE 4-1. Serial (left) and parallel (right) transfers of digital data.

A unit of data (e.g., a byte or 16-bit word) may be sent in *serial* form over a single wire by transferring each bit value in succession, as shown on the left in Figure 4-1. A much faster way of transferring the same unit of data is to use several wires, with each bit value being transferred over its own wire, as shown on the right in Fig. 4-1. This is called *parallel* data transfer. If eight wires are used for data transfer, a byte can be transferred in an eighth of the time required for serial transfer.

A group of wires used to transfer data between several devices is called a *data bus*. Each device connected to the bus is identified by an address or a range of addresses. Only one device at a time can transmit information on a bus, and in most cases only one device receives the information. The sending device transmits both the information and the address of the device that is intended to receive the information. Some buses use separate wires for the data and addresses, whereas others transmit both over the same wires. The width of a bus refers to the number of wires used to transmit data. For example, a 32-bit bus transmits 32 bits of data simultaneously. Buses usually have additional wires, such as a ground wire to establish a reference potential for the voltage signals on the other wires, wires to carry control signals, and wires at one or more constant voltages to provide electrical power to circuit boards attached to the bus.

4.2 ANALOG DATA AND CONVERSION BETWEEN ANALOG AND DIGITAL FORMS

Analog Representation of Data

Numerical data can be represented in electronic circuits by analog form instead of digital form by a voltage or voltage pulse whose amplitude is proportional to the number being represented, as shown in Fig. 4-2A. An example of analog representation is a voltage pulse produced by a photomultiplier tube attached to a scintillation detector. The amplitude (peak voltage) of the pulse is proportional to the amount of energy deposited in the detector by an x- or gamma ray. Another example is the signal from the video camera attached to the image intensifier tube of a fluoroscopy system; the voltage at each point in time is proportional to the intensity of the x-rays incident on a portion of the input phosphor of the image intensifier tube (Fig. 4-2C).

Advantages and Disadvantages of the Analog and Digital Forms

There is a major disadvantage to the electronic transmission of data in analog form—the signals become distorted. Causes of this distortion include inaccuracies

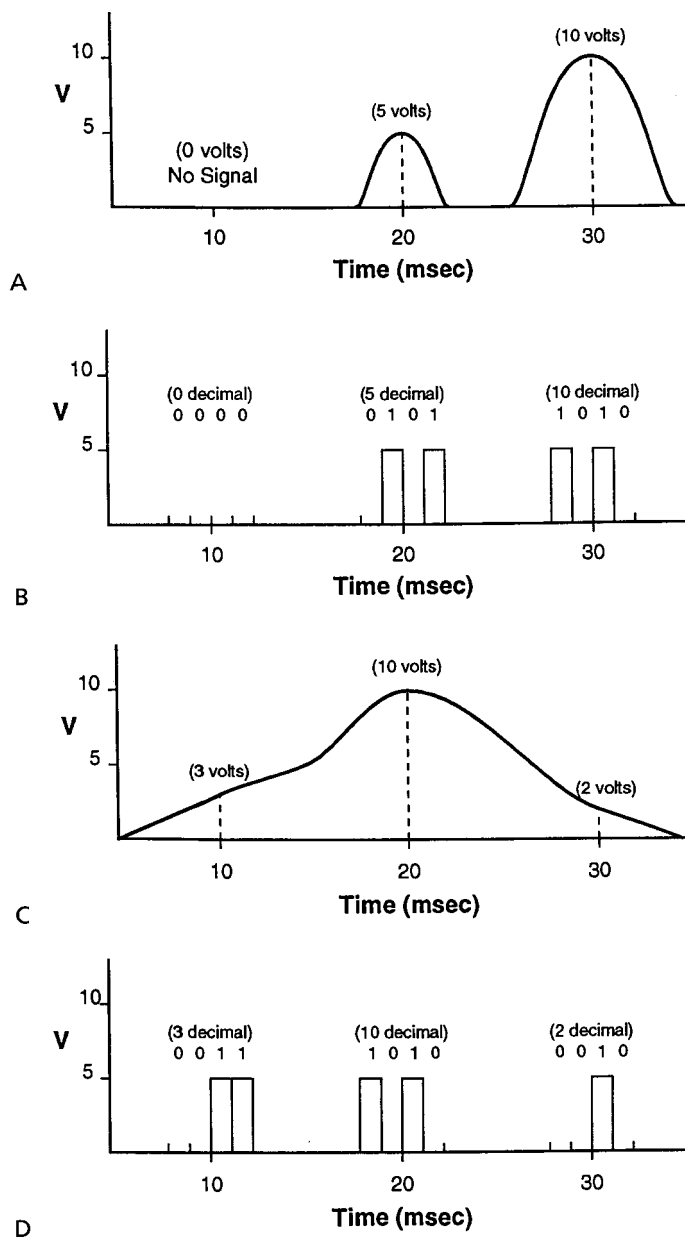


FIGURE 4-2. Analog and digital representation of numerical data. **A:** Three analog voltage pulses, similar to those produced by a photomultiplier tube attached to a scintillator. The height of each pulse represents a number. **B:** These same numbers represented in digital form. **C:** A continuously varying analog signal, such as that from the video camera in a fluoroscopy system. The height of the signal at each point in time represents a number. **D:** The values of this signal, sampled at three points, represented in digital form.

when signals are amplified, attenuation losses, and electronic noise—small stray voltages that exist on circuits and become superimposed on the data. The more the data are transferred, the more distorted they become. On the other hand, data stored or transferred in digital form are remarkably immune to the accumulation of error because of signal distortion. These distortions are seldom of sufficient amplitude to cause a 0 to be mistaken for a 1 or vice versa. Furthermore, most digital circuits do not amplify the incoming data, but make a fresh copy of it, thus preventing distortions from accumulating during multiple transfers.

The digital form facilitates other safeguards. When information integrity is critical, additional redundant information can be sent with each group of bits to permit the receiving device to detect errors or even correct them. A simple error detection method uses parity bits. An additional bit is transmitted with each group of bits, typically with each byte. The bit value designates whether an even or an odd number of bits were in the “1” state. The receiving device determines the parity and compares it with the received parity bit. If the parity of the data and the parity bit do not match, an odd number of bits in the group have errors.

Data can often be transferred in less time using the analog form. However, digital circuits are likely to be less expensive.

Conversion of Analog Data to Digital Form

The transducers, sensors, or detectors of most electronic measuring equipment, including medical imaging devices, produce analog data. If such data are to be analyzed by a computer or other digital equipment, they must be converted into digital form. Devices that perform this function are called *analog-to-digital converters* (ADCs). ADCs are essential components of multichannel analyzers, modern nuclear medicine scintillation cameras, computed radiography systems, modern ultrasound systems, MRI scanners, and CT scanners.

Most analog signals are continuous in time, meaning that at every point in time the signal has a value. However, it is not possible to convert the analog signal to a digital signal at every point in time. Instead, certain points in time must be selected at which the conversion is to be performed. This process is called *sampling*. Each analog sample is then converted into a digital signal. This conversion is called *digitization* or *quantization*.

An ADC is characterized by its *sampling rate* and the *number of bits of output* it provides. The sampling rate is the number of times a second that it can sample and digitize the input signal. Most radiologic applications require very high sampling rates. An ADC produces a digital signal of a fixed number of bits. For example, one may purchase an 8-bit ADC, a 10-bit ADC, or a 12-bit ADC. The number of bits of output is just the number of bits in the digital number produced each time the ADC samples and quantizes the input analog signal.

The digital representation of data is superior to analog representation in its resistance to the accumulation of errors. However, there are also disadvantages to digital representation, an important one being that *the conversion of an analog signal to digital form causes a loss of information*. This loss is due to both sampling and quantization. Because an ADC samples the input signal, the values of the analog signal between the moments of sampling are lost. If the sampling rate of the ADC is sufficiently rapid that the analog signal being digitized varies only slightly during the intervals between sampling, the sampling error will be small. There is a minimum sampling rate requirement, the *Nyquist limit* (discussed in Chapter 10) that ensures the accurate representation of a signal.

Quantization also causes a loss of information. An analog signal is continuous in magnitude, meaning that it may have any value between a minimum and maximum. For example, an analog voltage signal may be 1.0, 2.5, or 1.7893 V. In contrast, a digital signal is limited to a finite number of possible values, determined by the number of bits used for the signal. As was shown earlier in this chapter, a 1-bit digital signal is limited to two values, a 2-bit signal is limited to four values, and an

TABLE 4-5. MAXIMAL ERRORS WHEN DIFFERENT NUMBERS OF BITS ARE USED TO APPROXIMATE AN ANALOG SIGNAL

Number of Bits	Number of Values	Maximal Quantization Error (%)
1	2	25
2	4	12.5
3	8	6.2
8	256	0.20
12	4,096	0.012

N -bit signal is restricted to 2^N possible values. The quantization error is similar to the error introduced when a number is “rounded off.” Table 4-5 lists the maximal percent errors associated with digital signals of different numbers of bits.

There are additional sources of error in analog-to-digital conversion other than the sampling and quantization effects described above. For example, some averaging of the analog signal occurs at the time of sampling, and there are inaccuracies in the quantization process. In summary, a digital signal can only approximate the value of an analog signal, causing a loss of information during analog-to-digital conversion.

No analog signal is a perfect representation of the quantity being measured. Statistical effects in the measurement process and stray electronic voltages (“noise”) cause every analog signal to have some uncertainty associated with it. To convert an analog signal to digital form without a significant loss of information content, the ADC must sample at a sufficiently high rate and provide a sufficient number of bits so that the error is less than the uncertainty in the analog signal being digitized. In other words, an analog signal with a large *signal-to-noise ratio* (SNR) requires an ADC providing a large number of bits to avoid reducing the SNR.

Digital-to-Analog Conversion

It is often necessary to convert a digital signal to analog form. For example, to display digital images from a CT scanner on a video monitor, the image information must be converted from digital form to an analog voltage signal. This conversion is performed by a *digital-to-analog converter* (DAC). It is important to recognize that the information lost by analog-to-digital conversion is not restored by sending the signal through a DAC (Fig. 4-3).

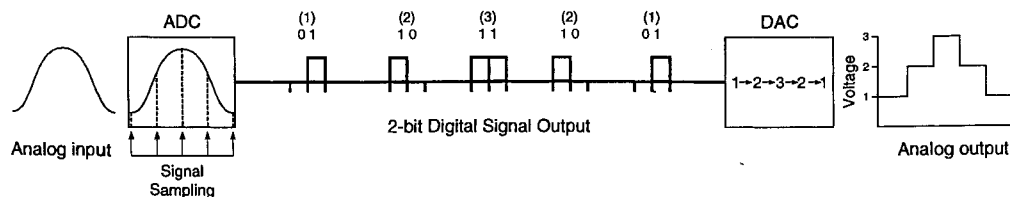


FIGURE 4-3. Analog-to-digital (ADC) conversion and digital-to-analog (DAC) conversion. In this figure, a 2-bit ADC samples the input signal five times. Note that the output signal from the DAC is only an approximation of the input signal to the ADC because the 2-bit digital numbers produced by the ADC can only approximate the continuously varying analog signal.

4.3 COMPONENTS AND OPERATION OF COMPUTERS

A computer consists of a central processing unit (CPU), main memory, and input/output (I/O) devices, all linked together by one or more data pathways called data buses (Fig. 4-4). The main memory of the computer stores the program (sequence of instructions) being executed and the data being processed. The CPU executes the instructions in the program to process the data.

I/O devices enable information to be entered into and retrieved from the computer and to be stored. The I/O devices of a computer usually include a keyboard, a mouse or other pointing device, a video interface and video monitor, several mass storage devices, and often a printer. The keyboard, pointing device, video interface, and video monitor enable the operator to communicate with the computer. Mass storage devices permit the storage of large volumes of programs and data. They include floppy magnetic disk drives, fixed magnetic disk drives, optical disk drives, and magnetic tape drives.

When the CPU or another device sends data to memory or a mass storage device, it is said to be *writing* data, and when it requests data stored in memory or on a storage device, it is said to be *reading* data. Computer memory and data storage devices permit either random access or sequential access to the data. The term *random access* describes a storage medium in which the locations of the data may be read or written in any order, and *sequential access* describes a medium in which data storage locations can only be accessed in a serial manner. Most solid-state memories, magnetic disks, and optical disks are random access, whereas magnetic tape permits only sequential access.

Main Memory

Main memory provides temporary storage for the instructions (computer program) currently being executed and the data currently being processed by the computer.

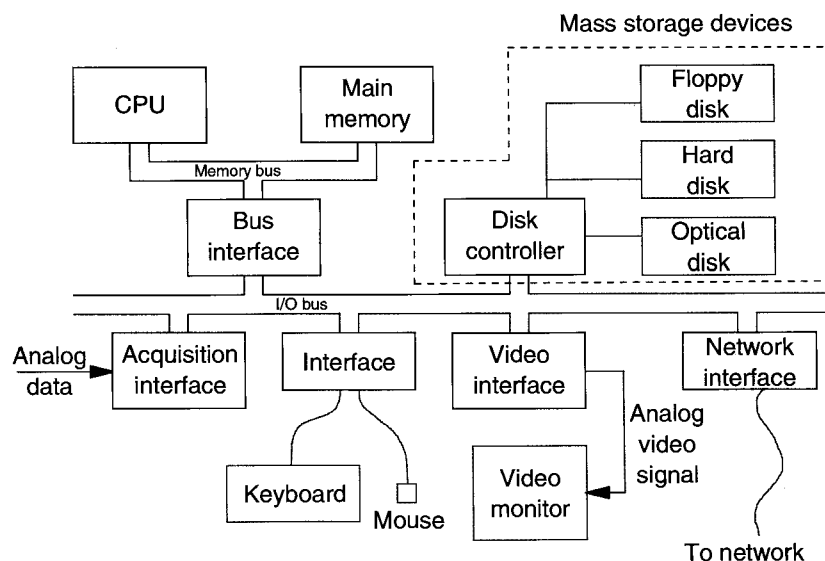


FIGURE 4-4. Components of a computer.

Main memory is used for these functions instead of mass storage devices because the data transfer rate between the CPU and main memory is much faster than that between the CPU and the mass storage devices.

Main memory consists of a large number of data storage locations, each of which contains the same number of bits (usually 1 byte). A unique number, called a *memory address*, identifies each storage location. Memory addresses usually start at zero and increase by one to one less than the total number of memory locations, as shown in Fig. 4-5. A memory address should not be confused with the data stored at that address.

When a CPU performs a memory write, it sends both a memory address, designating the desired storage location, and the data to the memory. When it performs a memory read, it sends the address of the desired data to the memory and, in return, is sent the data at that location.

One type of memory is commonly called *random access memory* (RAM). In common usage, RAM refers to memory onto which the CPU can both read and write data. A better name for such memory is read-write memory. A disadvantage of most read-write memory is that it is *volatile*, meaning that data stored in it are lost when electrical power is turned off.

Another type of memory is called *read-only memory* (ROM). The CPU can only read data from ROM; it cannot write or erase data on ROM. The main advantage to ROM is that data stored in it are not lost when power is lost to the computer. ROM is usually employed to store frequently used programs provided by the manufacturer that perform important functions such as setting up the computer for operation when power is turned on. Although the term *RAM* is commonly used to distinguish read-write memory from ROM, ROM does permit random access.

The size of the main memory can affect the speed at which a program is executed. A very large program or one that processes large volumes of data will usually be executed more quickly on a computer with a large memory. A computer with a small

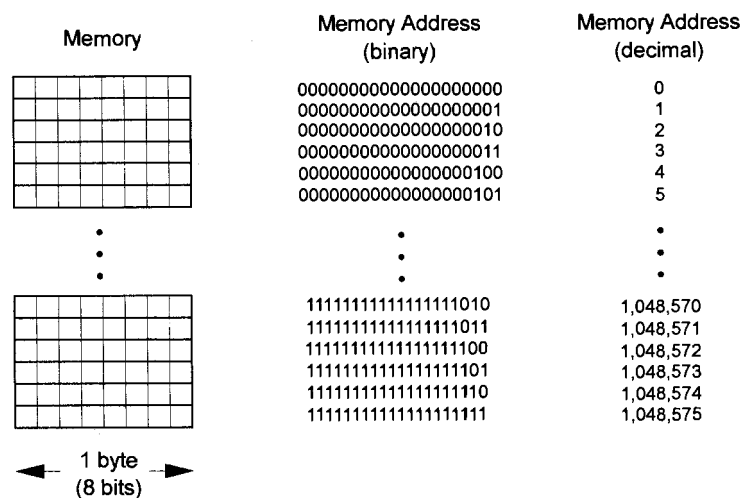


FIGURE 4-5. Main memory, containing 1 megabyte (2^{20} bytes). Each byte in memory is identified by a unique memory address, just as each mailbox in a city is identified by a postal address.

memory can fit only small portions of the program and data into its memory at one time, and must frequently interrupt the processing of data to load other portions of the program and data into the memory from the magnetic disk or other storage devices.

It would be very expensive to have a large main memory that could provide program instructions and data to the CPU as fast as the CPU can process them. The main memories of most computers today use a relatively slow but inexpensive type of memory called dynamic RAM (DRAM). DRAM uses one capacitor (an electrical charge storage device) and one transistor for each bit. It is called dynamic because each bit must be read and refreshed several times a second to maintain the data. To compensate for the low speed of DRAM, many computers contain one or more smaller but faster memories, called *cache memories*, between the main memory and the CPU. They maintain exact copies of small portions of the program and data in the main memory. A cache memory usually consists of static RAM (SRAM), which uses several transistors to represent each bit. SRAM stores data while power is on without requiring refreshing and can be read or written several times faster than DRAM, but requires more space and is more expensive. Cache memories are effective because program segments and data that were recently read from the memory by the CPU are very likely to be read again soon.

Computer Program and Central Processing Unit

A computer program is a sequence of instructions and the central processing unit (CPU) is a set of electronic circuits that executes them. A CPU fetches and executes the instructions in a program sequentially; it fetches an instruction from main memory, executes that instruction, and then fetches the next instruction in the program sequence. A CPU contained in a single computer chip is called a *microprocessor*.

The CPU contains a small number of data storage locations, called *storage registers*, for storing data and memory addresses. (For example, one computer system has 16 registers, each containing 32 bits.) Operations can be performed on data in the registers much more quickly than on data in memory. The CPU also contains an *arithmetic logic unit* (ALU) that performs mathematical operations and comparisons between data stored in the registers or memory. In addition, the CPU contains a clock. This clock is not a time-of-day clock (although the computer may have one of these as well), but a circuit that produces a periodic stream of logic pulses that serve as timing marks to synchronize the operation of the various components of the computer. CPU clock speeds are usually described in millions or billions of hertz (Hz); an Hz is defined as one cycle per second.

An instruction can cause the CPU to perform one or a combination of four actions:

1. Transfer a unit of data (e.g., a 16- or 32-bit word) from a memory address, CPU storage register, or I/O device to another memory address, register, or I/O device.
2. Perform a mathematical operation between two numbers in storage registers in the CPU or in memory.
3. Compare two numbers or other pieces of data in registers or in memory.
4. Change the address of the next instruction in the program to be executed to another location in the program. This type of instruction is called a *branching instruction*.

A CPU fetches and executes the instructions in a program sequentially until it fetches a branching instruction. The CPU then jumps to a new location in the program that is specified by the branching instruction and, starting at the new location, again executes instructions sequentially. There are two types of branching instructions, unconditional and conditional. *Conditional branching instructions* cause a branch to occur only if a specified condition is met. For example, a conditional branching instruction may specify that a branch is to occur only if the number stored in a particular CPU storage register is zero. Conditional branching instructions are especially important because they permit the computer to make “decisions” by allowing the data to alter the order of instruction execution.

Each instruction in a program has two parts. The first part specifies the operation, such as a data transfer, an addition of two numbers, or a conditional branch, to be performed. The second part describes the location of the data to be operated on or compared and the location where the result is to be stored. In the case of an addition, it contains the locations of the numbers to be added and the location where the sum is to be stored. In the case of a branching instruction, the second part is the memory address of the next instruction to be fetched and executed.

The CPU is a bottleneck in a conventional computer. A computer may have many megabytes of data to be processed and a program requiring the execution of instructions many billions of times. However, the CPU executes the instructions only one or a few at a time. One way to increase speed is to build the circuits of the CPU and memory with faster components so the computer can run at a higher clock speed. This makes the computer more expensive and there are practical and technical limits to how fast a circuit can operate.

The other method of relieving the bottleneck is to design the CPU to perform *parallel processing*. This means that the CPU performs some tasks simultaneously rather than sequentially. Most modern CPUs incorporate limited forms of parallel processing. Some specialized CPUs, such as those in array processors (to be described later in this chapter), rely heavily on parallel processing. One form of parallel processing is pipelining. Several steps, each requiring one clock cycle, are required for a CPU to execute a single instruction. Older CPUs, such as the Intel 8088, 80286, and 80386 microprocessors, completed only one instruction every few cycles. The instruction execution circuits of newer microprocessors are organized into pipelines, a pipeline having several stages that operate simultaneously. Instructions enter the pipeline one at a time, and are passed from one stage to the next. Several instructions are in different stages of execution at the same time. It takes a several clock cycles to fill the pipeline, but, once full, it completes the execution of one instruction per cycle. Pipelining permits a CPU to attain an average instruction execution rate of almost one instruction per cycle. (Pipelining does not yield exactly one instruction per cycle because some instructions interfere with its operation, in particular ones that modify the data to be operated on by immediately subsequent instructions, and branching instructions, because they change the order of instruction execution.) Superscalar CPU architectures incorporate several pipelines in parallel and can attain average instruction execution rates of two to five instructions per CPU clock cycle.

Input/Output Bus and Expansion Slots

Data buses were described previously in this chapter. The input/output buses of most computers are provided with several expansion slots into which printed circuit

boards can be plugged. The boards installed in these slots can include modems, cards to connect the computer to a computer network, controllers for mass storage devices, display interface cards to provide a video signal to a video monitor, sound cards to provide audio signals to speakers, and acquisition interfaces to other devices such as scintillation cameras. The provision of expansion slots on this bus makes it possible to customize a general-purpose computer for a specific application and to add additional functions and capabilities.

Mass Storage Devices

Mass storage devices permit the nonvolatile (i.e., data are not lost when power is turned off) storage of programs and data. Mass storage devices include floppy disk drives, hard disk drives (nonremovable hard disks are called fixed disks), magnetic tape drives, and optical (laser) disk units. Each of these devices consists of a mechanical drive; the storage medium, which may be removable; and an electronic controller. Despite the wide variety of storage media, there is not one best medium for all purposes, because they differ in storage capacity, data access time, data transfer rate, cost, and other factors.

Magnetic disks are spinning disks coated with a material that may be readily magnetized. Just above the spinning disk is a read/write head that, to read data, senses the magnetization of individual locations on the disk and, to write data, changes the direction of the magnetization of individual locations on the disk. A disk drive has a read/write head on each side of a platter so that both sides can be used for data storage. The data is stored on the disk in concentric rings called *tracks*. The read/write heads move radially over the disk to access data on different tracks. The *access time* of a disk is the time required for the read/write head to reach the proper track (head seek time) and for the spinning of the disk to bring the data of interest to the head (rotational latency). The *data transfer rate* is the rate at which data are read from or written to the disk once the head and disk are in the proper orientation; it is primarily determined by the rotational speed of the disk and the density of data storage in a track.

A typical hard disk drive, as shown in Fig. 4-6, has several rigid platters stacked above each other on a common spindle. The platters continuously rotate at a high speed (typically 5,400 to 15,000 rpm). The read/write heads aerodynamically float microns above the disk surfaces on air currents generated by the spinning platters. A spinning hard disk drive should not be jarred because it might cause a “head crash” in which the head strikes the disk, gouging the disk surface and destroying the head, with a concomitant loss of data. Disks must also be kept clean; a hair or speck of dust can damage the disk and head. For this reason, the portion containing the disks and read/write heads is usually sealed to keep dirt out. Hard disk drives with removable disk platters are less common. The storage capacities of hard disks are very large, ranging from approximately 10 to 180 gigabytes. Although their access times and data transfer rates are very slow compared to memory, they are much faster than those of floppy disks or magnetic tape. For these reasons, hard disks are used on most computers to store frequently used programs and data.

Floppy disks are removable plastic disks coated with a magnetizable material. They are operationally similar to hard disks, but their rotation is intermittent (spinning only during data access), they spin more slowly, and the read/write heads maintain physical contact with the disk surfaces. Their access times are considerably longer and their data transfer rates are much slower than those of hard disks. Their storage capacities, typi-



FIGURE 4-6. Hard-disk drive. A read/write head is visible at the end of the arm overlying the top disk platter. (Photograph courtesy of Seagate Technology, Inc.)

cally about 1.4 megabytes, are much smaller. They are commonly used for sending programs and data from one site to another and for making “backup” copies of important programs and data. Floppy disk drives with capacities of 100 megabytes or more have been developed. However, there are not yet standards for them, and in general those from one manufacturer cannot be read by disk drives of other manufacturers.

Magnetic tape is plastic tape coated with a magnetizable substance. Its average data access times are very long, because the tape must be read serially from the beginning to locate a particular item of data. There are several competing tape-cartridge formats available today. A single tape cartridge can store a very large amount of data, in the range of 4 to 110 gigabytes. Common uses are to “back up” (make a copy for the purpose of safety) large amounts of important information and for archival storage of digital images.

An optical disk is a removable disk that rotates during data access and from which data are read and written using a laser. There are three categories of optical disks—read-only; write-once, read-many-times (WORM); and rewritable. Read-only disks are encoded with data that cannot be modified. To read data, the laser illuminates a point on the surface of the spinning disk and a photodetector senses the intensity of the reflected light. CD-ROMs are an example of read-only optical disks. Most CD-ROMs have capacities of 650 megabytes. CD-ROMs are commonly used by vendors to distribute commercial software and music. WORM devices permit any portion of the disk surface to be written upon only once, but to be read many times. This largely limits the use of WORM optical disks to archival storage. To store data, a laser at a high intensity burns small holes in a layer of the disk. To read data, the laser, at a lower intensity, illuminates a point on the spinning disk and a photodetector senses the intensity of the reflected light. CD-Rs (for recordable) are a common type of WORM disk. Rewritable optical disks permit the stored data to be changed many times. There are currently two types of rewritable optical disks—phase-change and magneto-

optical. The recording material of a phase-change disk has the property that, if heated to one temperature and allowed to cool, it becomes crystalline, whereas if heated to another temperature, it cools to an amorphous phase. To write data on a phase-change disk, the laser heats a point on the recording film, changing its phase from crystalline to amorphous or vice versa. The transparency of the amorphous material differs from that in the crystalline phase. Above the layer of recording material is a reflective layer. Data are read as described above for WORM disks. A magneto-optical disk contains a layer of recording material whose magnetization may be changed by a nearby magnet if heated above a temperature called the *Curie temperature*. Writing data on a magneto-optical disk requires a small electromagnet above the disk and a laser below. The laser heats a small domain on the layer of recording material above the Curie temperature, causing its magnetization to align with that of the electromagnet. When the domain cools below the Curie temperature, the magnetism of that domain becomes “frozen.” The magneto-optical disk is read by the laser at a lower intensity. The polarization of the reflected light varies with the magnetization of each domain on the disk. CD-RWs are a common type of rewritable optical disk. The CD is likely to be displaced by a new standard for optical disks, the DVD, that provides a much larger storage capacity. DVDs are available in read-only, recordable, and rewritable forms. (“DVD-RAM” is one of two competing formats for rewritable DVDs.) Advantages of optical disks include large storage capacity, on the order of 650 megabytes to several gigabytes, and much faster data access than magnetic tape. Optical disks provide superior data security than most other media because they are not subject to head crashes, as are magnetic disks; are not as vulnerable to wear and damage as magnetic tape; and are not vulnerable to magnetic fields.

TABLE 4-6. COMPARISON OF CHARACTERISTICS OF MASS STORAGE DEVICES AND MEMORY^a

	Removable	Storage Capacity	Access Time (Average)	Transfer Rate	Cost per Disk/Tape	Media Cost per GB
Floppy disk	Yes	1.2 to 1.4 MB	300 msec	0.03 MB/sec	\$0.30	\$210
Hard disk	Usually not	10 to 182 GB	6 to 15 msec	3 to 50 MB/sec	NA	
Optical disk, CD-ROM, CD-R, CD-RW	Yes	650–700 MB	100 to 150 msec (for 24×)	3.6 MB/sec (for 24×)	\$0.50 (CD-R) \$1.10 (CD-RW)	\$0.80 (CD-R) \$1.70 (CD-RW)
Optical disk, DVD-ROM, DVD-R, DVD-RAM	Yes	4.7 to 17 GB 3.9 GB 5.2 GB		8.1 MB/s	\$24 (DVD-RAM)	\$4.60 (DVD-RAM)
Magnetic tape (cartridge)	Yes	45 MB to 110 GB	Seconds to minutes	0.125 to 10 MB/sec	\$150 (110 GB super DLT)	\$1.35 (110 GB super DLT)
Solid-state memory	No	64 MB to 1.5 GB	1 to 80 msec	24 to 250 MB/sec	NA	

^aValues are typical for 2001-vintage computers. Cost refers to one diskette or tape cartridge. Storage capacities and transfer rates are for uncompressed data; they would be higher for compressed data. msec, milliseconds; nsec, nanoseconds (10⁶ nsec = 1 msec); MB, megabytes; GB, gigabytes; DLT, digital linear tape.

Table 4-6 compares the characteristics of mass storage devices and memory. Because there is no best device for all purposes, most computer systems today have at least a hard disk drive, a floppy disk drive, and a drive capable of reading optical disks.

Display Interface

An important use of a computer is to display information in visual form. The information may be displayed on a video monitor using a cathode ray tube (CRT) or a flat-panel display. The information may also be presented in hard-copy form, most commonly using photographic film or paper.

Medical imaging computers have video interfaces and display devices, both to assist the user in controlling the computer and to display information. The *video interface* is a device that receives digital data from the memory of a computer via the computer's bus and, using one or more digital-to-analog converters, converts it into an analog video signal for display on a video monitor. Most video interfaces contain their own memory for storing displayed images, although some computers use a portion of main memory for this purpose. The video interface and display devices will be described in greater detail below (see Display).

Keyboard and Pointing Device

Usually, a computer is equipped with keyboard and a pointing device, such as a mouse, trackball, or joystick, to enable the user to direct its operations and to enter data into it.

Acquisition Interface

Until recently, many medical imaging devices, such as scintillation cameras, sent analog data directly to attached computers. In most cases today, the analog data are converted into digital form by the imaging device, which sends the digital data to the computer. A computer's acquisition interface accepts data from an attached device and transfers it to memory via the computer's bus. An acquisition interface that accepts analog data first converts it into digital form using one or more ADCs. Some acquisition interfaces store the received information in their own dedicated memories to permit the computer to perform other tasks during data acquisition. An imaging computer that is not directly connected to an imaging device may instead receive images over a computer network, as described below.

Communications Interface

Computers serve not only as data processing and display devices, but also as communications devices. Thus many, if not most, computers are connected, either intermittently or continuously, to other computers or a computer network. A modem (contraction of "modulator-demodulator") is a device to permit a computer to transfer information to another computer or to a network by analog telephone lines. The modem converts the digital signals from a computer to an analog signal and the modem on the other end converts the analog signal back to digital form. The analog form used by modems is different from the analog signals discussed previously in this chapter, in which the information was encoded as the amplitude of the signal. Instead, modems produce a sinusoidal signal of constant amplitude and encode the information by varying the frequency of the signal (frequency modulation).

A computer may also be directly connected to a network using a network interface card. The card converts digital signals from the computer into the digital form

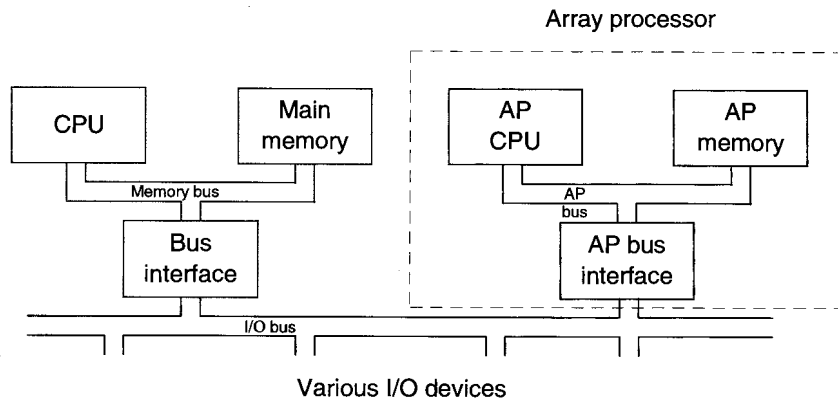


FIGURE 4-7. Computer with an array processor.

required by the network and vice versa. The transfer of information using computer networks is discussed in detail in Chapter 17.

Array Processor

The speed of a general-purpose computer in performing mathematical operations, especially on floating-point numbers, can be greatly enhanced by adding an *array processor*. An array processor is a specialized computer designed to perform the mathematical manipulation of data very quickly. It consists of a CPU, a small high-speed memory, a high-speed bus, and an interface to connect it to the bus of the main computer. The main computer is called the *host computer*. An array processor's CPU relies heavily on parallel processing for its speed. Parallel processing is particularly efficient for processing images. The host computer sends the data to be processed from its memory to the array processor's memory along with the instructions to execute in processing the data. The array processor processes the data and returns the processed data to the memory of the host computer. Figure 4-7 is a diagram of a computer with an array processor.

4.4 PERFORMANCE OF COMPUTER SYSTEMS

There are many factors that affect the time required for a computer to complete a task. These include the speed of the CPU (as reflected its clock speed), how extensively the CPU uses parallel processing, width and the clock speed of the data bus between the CPU and main memory, the size of the memory, and the access times and data transfer rates of mass storage devices. A larger memory permits more of the program being executed and the data being processed to reside in memory, thereby reducing the time that the CPU is idle while awaiting the transfer of program segments and data from mass storage devices into memory. The widths of the data buses affect the speed of the computer; a 32-bit bus can transfer twice as much data per clock cycle as a 16-bit bus.

There are several ways to compare the performance of computer systems. One indicator of CPU speed is the speed of its clock. If two computers are oth-

erwise identical, the computer with the faster clock speed is the fastest. For example, if the CPUs of two computers are both Intel Pentium microprocessors, the CPU running at 1.2 gigahertz (GHz) is faster than the CPU running at 800 megahertz (MHz).

The clock speed of a CPU is not as useful in comparing computers with different CPUs, because one CPU architecture may execute more instructions per clock cycle than another. To avoid this problem, the speed of a CPU may be measured in *millions of instructions per second* (MIPS). If a computer is used to perform extensive mathematical operations, it may be better to measure its performance in millions or billions of floating-point operations per second (megaflops or gigaflops). Another way to compare the performance of computers is to compare the times it takes them to execute standard programs. Programs used for this purpose are called *benchmarks*.

The ability of a computer to execute a particular program in less time than another computer or to perform a greater number of MIPS or megaflops does not prove that it is faster for all applications. For example, the execution speed of one application may rely primarily on CPU speed and CPU-memory transfer rates, whereas another application's speed of execution may largely depend on the rate of transfer of data between the hard disk and main memory. When evaluating computers for a particular purpose, such as reconstructing tomographic images, it is better to compare their speed in performing the actual tasks expected of them than to compare them solely on the basis of clock speed, MIPS, megaflops, or benchmark performance.

4.5 COMPUTER SOFTWARE

Computer Languages

Machine Language

The actual binary instructions that are executed by the CPU are called *machine language*. There are disadvantages to writing a program in machine language. The programmer must have a very detailed knowledge of the computer's construction to write such a program. Also, machine-language programs are very difficult to read and understand. For these reasons, programs are almost never written in machine language.

High-Level Languages

To make programming easier and to make programs more readable, *high-level languages*, such as Fortran, Basic, Pascal, C, and Java, have been developed. These languages enable computer programs to be written using statements similar to simple English, resulting in programs that are much easier to read than those in machine language. Also, much less knowledge of the particular computer is required to write a program in a high-level language. Using a program written in a high-level language requires a program called a *compiler* or an *interpreter* to convert the statements written in the high-level language into machine-language instructions that the CPU can execute. Figure 4-8 shows a simple program in the language Basic.


```
10 I = 1
20 PRINT I
30 I = I+1
40 IF I<11 THEN GOTO 20
50 END
```

FIGURE 4-8. A simple program in the language Basic. The program prints the integers from 1 to 10 on the video monitor.

Applications Programs

An *applications program* is a program that performs a specific function, such as word processing, accounting, scheduling, or image processing. It may be written in machine language or a high-level language. A good applications program is said to be *user-friendly*, meaning that only a minimal knowledge of the computer is needed to use the program.

Operating System

Even when a computer is not executing an assigned program and seems to be idle, there is a fundamental program, called the *operating system*, running. When the user instructs a computer to run a particular program, the operating system copies the program into memory from a disk, transfers control to it, and regains control of the computer when the program has finished. The operating system also handles most of the complex aspects of input/output operations and controls the sharing of resources on multiuser computers. Common operating systems include Windows, Windows NT, and Linux for IBM-compatible PCs; the Mac OS on Apple computers; and Unix for workstations.

Most operating systems store information on mass storage devices as *files*. A single file may contain a program, one or more images, patient data, or a manuscript. Each mass storage device has a *directory* that lists all files on the device and describes the location of each.

The operating system can be used to perform a number of utility functions. These utility functions include listing the files on a mass storage device, determining how much unused space remains on a mass storage device, deleting files, or making a copy of a file on another device. To perform one of these functions, the user types a command or designates a symbol on the computer screen using a pointing device, and the operating system carries it out.

Computer Security

The goals of computer security are to deny unauthorized persons access to confidential information, such as patient data, and to protect software and data from accidental or deliberate modification or loss. A major danger in using computers is that the operating system, applications programs, and important data are often stored on a single disk; an accident could cause all of them to be lost. The primary threats to data and software on computers are mechanical or electrical mal-

function, such as a disk head crash; human error, such as the accidental deletion of a file or the accidental formatting of a disk (causing the loss of all information on the disk); and malicious damage. To reduce the risk of data loss, important files should be copied ("backed up") onto floppy disks, optical disks, or magnetic tape at regularly scheduled times.

Programs written with malicious intent are a threat to computers. The most prevalent of these is the computer virus. A virus is a string of instructions hidden in a program. If a program containing a virus is loaded into a computer and executed, the virus copies itself into other programs stored on mass storage devices. If a copy of an infected program is sent to another computer and executed, that computer becomes infected. Although a virus need not do harm, some deliberately cause damage, such as the deletion of all files on the disk on Friday the 13th. Viruses that are not intended to cause damage may interfere with the operation of the computer or cause damage because they are poorly written. There are three types of viruses—executable file infectors, boot-sector viruses, and macroviruses. Executable file infectors infect files containing executable instructions. An infected file must be imported and executed to trigger one of these. Boot-sector viruses infect the boot sectors of floppy disks. A boot sector is a small portion of a disk that contains instructions to be executed by the computer when power is first turned on. Turning power on to an IBM-compatible personal computer, if an infected floppy is in the floppy drive, will cause an infection. Macroviruses infect macros, programs attached to some word processing and other files. A computer cannot be infected with a virus by the importation of data alone or by the user reading an e-mail message. However, a computer can become infected if an infected file is attached to an e-mail message and if that file is executed.

There are other types of malicious programs. These include Trojan horses, programs that appear to serve one function, but have a hidden purpose; worms, programs that automatically spread over computer networks; and password grabbers, programs that store the passwords of persons logging onto computers for use by an unauthorized person. The primary way to reduce the chance of a virus infection or other problem from malicious software is to establish a policy forbidding the loading of storage media and software from untrustworthy sources. Commercial virus-protection software, which searches files on storage devices and files received over a network for known viruses and removes them, should be used. The final line of defense, however, is the saving of backup copies. Once a computer is infected, it may be necessary to reformat all disks and reload all software and data from the backup copies.

More sophisticated computer operating systems, such as Unix, provide security features including password protection and the ability to grant individual users different levels of access to stored files. Measures should be taken to deny unauthorized users access to your system. Passwords should be used to deny access, directly and over a network or a modem. Computers in nonsecure areas should be "logged off" when not in use. Each user should be granted only the privileges required to accomplish needed tasks. For example, technologists who acquire and process studies and interpreting physicians should not be granted the ability to delete or modify system software files or patient studies, whereas the system manager must be granted full privileges to all files on the system.

4.6 STORAGE, PROCESSING, AND DISPLAY OF DIGITAL IMAGES

Storage

A digital image is a rectangular, usually square, array of numbers. The portion of the image represented by a single number is called a *pixel* (for picture element). For example, in nuclear medicine planar images, a pixel contains the number of counts detected by the corresponding portion of the crystal of a scintillation camera. Figure 4-9 shows

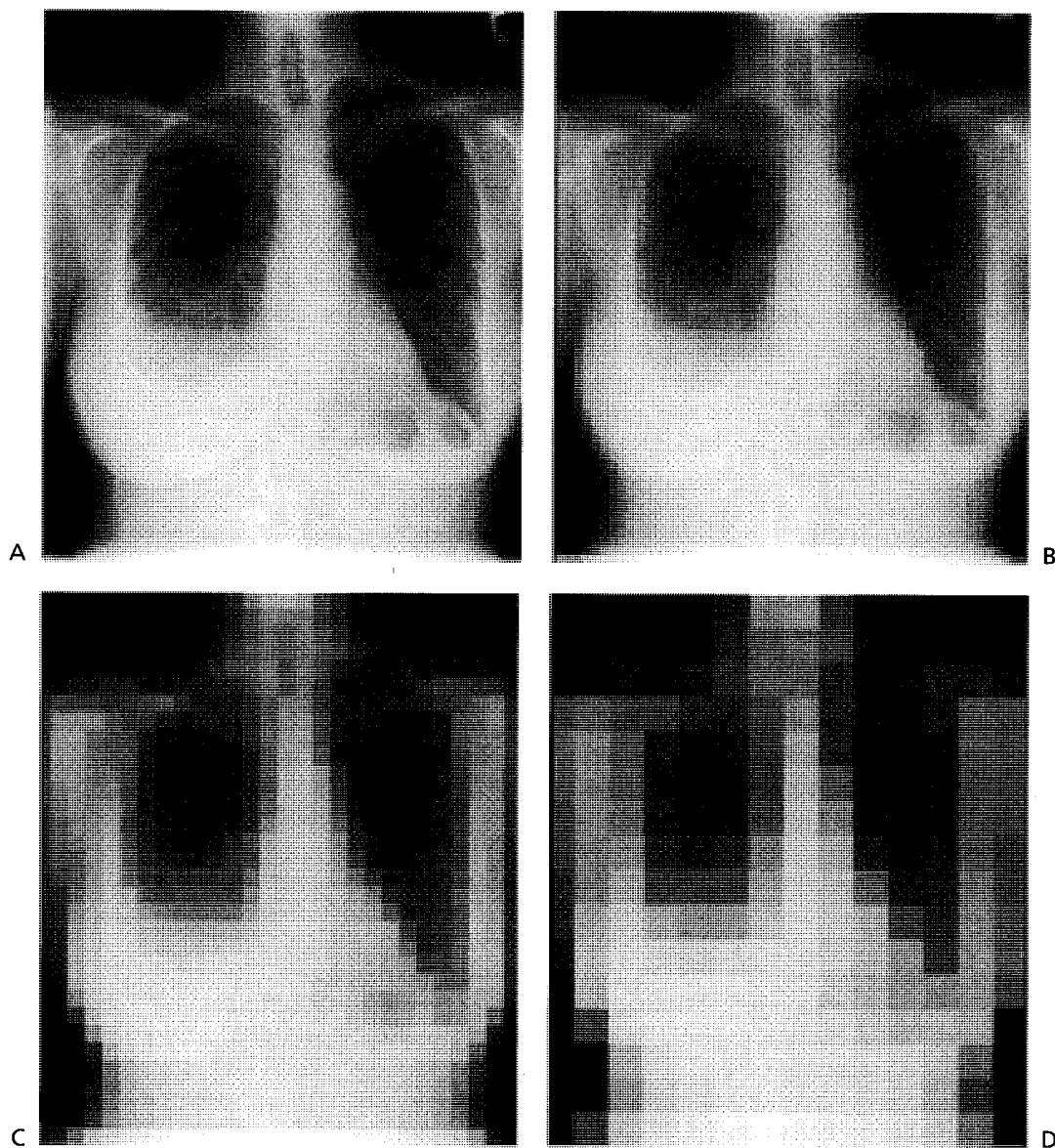


FIGURE 4-9. Effect of pixel size on image quality—digital chest image in formats of $1,024^2$, 64^2 , 32^2 , and 16^2 pixels (A, B, C, and D, respectively).

TABLE 4-7. TYPICAL RADIOLOGIC IMAGE FORMATS

Modality	Pixel Format	Bits per Pixel
Scintillation camera planar	64^2 or 128^2	8 or 16
SPECT	64^2 or 128^2	8 or 16
PET	128^2	16
Digital fluoroscopy, cardiac catheter lab	512^2 or 1024^2	8 to 12
Computed radiography, digitized chest films	Typically $2,000 \times 2,500$	10–12
Mammography (18×24 cm) or (24×30 cm)	Typically $1,800 \times 2,300$ to $4,800 \times 6,000$	12–16
X-ray CT	512^2	12
MRI	64^2 to $1,024^2$	12
Ultrasound	512^2	8

CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single photon emission computed tomography.

a digital image in four different pixel formats. The number in each pixel is converted into a visible light intensity when the image is displayed on a video monitor. Typical image formats used in radiology are listed in Table 4-7.

Imaging modalities with higher spatial resolution require more pixels per image so the image format does not degrade the resolution. In general, an image format should be selected so that the pixel size is on the order of the size of the smallest object to be seen. In the case of a fluoroscope with a 23 cm field of view, the 512^2 pixel format would be adequate for detecting objects as small as about a half a millimeter in size. To detect objects half this size, a larger format, such as 1024 by 1024 , should be selected. When it is necessary to depict the shape of an object, such as a microcalcification in x-ray mammography, an image format much larger than that needed to merely detect the object is required. Figure 4-9 shows the degradation of spatial resolution caused by using too small an image format. The penalty for using larger formats is increased storage and processing requirements and slower transfer of images.

The largest number that can be stored in a single pixel is determined by the number of bits or bytes used for each pixel. If 1 byte (8 bits) is used, the maximal number that can be stored in one pixel is 255 ($2^8 - 1$). If 2 bytes (16 bits) are used, the maximal number that can be stored is 65,535 ($2^{16} - 1$). The amount of contrast resolution provided by an imaging modality determines the number of bits required per pixel. Therefore, imaging modalities with higher contrast resolution require more bits per pixel. For example, the limited contrast resolution of ultrasound usually requires only 6 or 7 bits, and so 8 bits are commonly used for each pixel. On the other hand, x-ray CT provides high contrast resolution and 12 bits are required to represent the full range of CT numbers. Figure 4-10 shows the degradation of spatial resolution caused by using too few bits per pixel.

Pixel size is determined by dividing the distance between two points in the subject being imaged by the number of pixels between these two points in the image. It is approximately equal to the field of view of the imaging device divided by the number of pixels across the image. For example, if a fluoroscope has a 23-cm (9-inch) field of view and the images are acquired in a 512 -by- 512 format, then the approximate size of a pixel is $23 \text{ cm} / 512 = 0.45 \text{ mm}$, for objects at the face of the image intensifier.

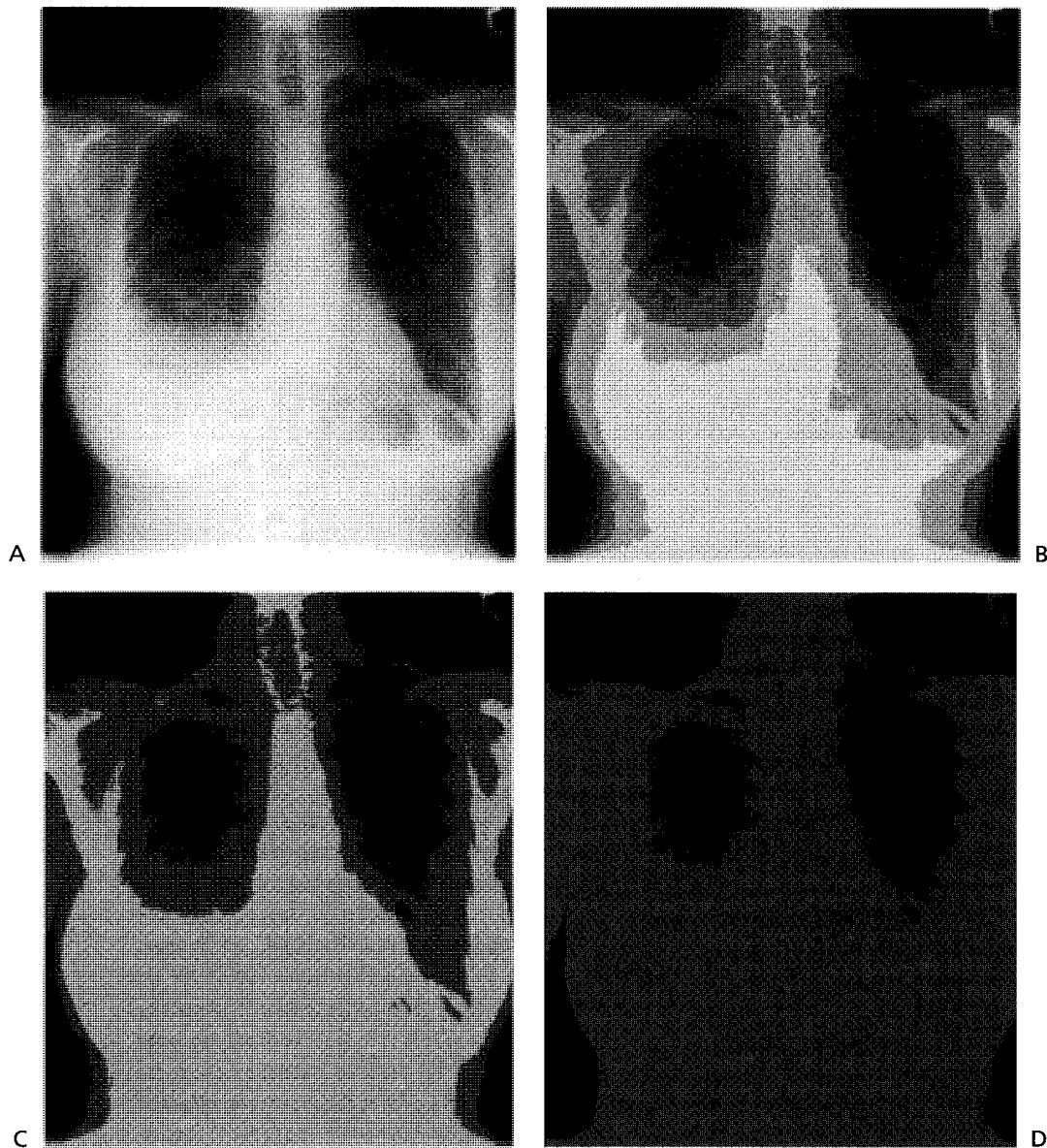


FIGURE 4-10. Effect of number of bits per pixel on image quality—digital chest image in formats of 8, 3, 2, and 1 bits per pixel (**A**, **B**, **C**, and **D**, respectively). Too few bits per pixel not only causes loss of contrast resolution, but also creates the appearance of false contours.

The total number of bytes required to store an image is the number of pixels in the image multiplied by the number of bytes per pixel. For example, the number of kilobytes required to store a 512-by-512 pixel image, if one byte is used per pixel, is

$$(512 \times 512 \text{ pixels}) (1 \text{ byte/pixel}) / (1,024 \text{ bytes/kB}) = 256 \text{ kB.}$$

Similarly, the number of 64-by-64 images that may be stored on a 1.4-megabyte floppy disk, if 16 bits are used per pixel, is

$$(1.4 \text{ MB/disk})(1,024^2 \text{ bytes/MB})/[(64 \times 64 \text{ pixels/image}) (2 \text{ bytes/pixel})] = 179 \text{ images/disk}$$

(if no other information is stored on the disk).

Processing

An important use of computers in medical imaging is to process digital images to display the information contained in the images in more useful forms. Digital image processing cannot add information to images. For example, if an imaging modality cannot resolve objects less than a particular size, image processing cannot cause such objects to be seen.

In some cases, the information of interest is visible in the unprocessed images and is merely made more conspicuous. In other cases, information of interest may not be visible at all in the unprocessed image data. X-ray computed tomography is an example of the latter case; observation of the raw projection data fails to reveal many of the structures visible in the processed cross-sectional images. The following examples of image processing are described only superficially and will be discussed in detail in later chapters.

The addition or subtraction of digital images is used in several imaging modalities. Both of these operations require the images being added or subtracted to be of the same format (same number of pixels along each axis of the images) and produce a resultant image (sum image or difference image) in that format. In image subtraction, each pixel in one image is subtracted from the corresponding image in a second image to yield the corresponding pixel value in the difference image. Image addition is similar, but with pixel-by-pixel addition instead of subtraction. Image subtraction is used in digital subtraction angiography to remove the effects of anatomic structures not of interest from the images of contrast-enhanced blood vessels (see Chapter 12) and in nuclear gated cardiac blood pool imaging to yield difference images depicting ventricular stroke-volume and ventricular wall dyskinesis (Chapter 21).

Spatial filtering is used in many types of medical imaging. Medical images often have a grainy appearance, called *quantum mottle*, caused by the statistical nature of the acquisition process. The visibility of quantum mottle can be reduced by a spatial filtering operation called *smoothing*. In most spatial smoothing algorithms, each pixel value in the smoothed image is obtained by a weighted averaging of the corresponding pixel in the unprocessed image with its neighbors. Although smoothing reduces quantum mottle, it also blurs the image. Images must not be smoothed to the extent that clinically significant detail is lost. Spatial filtering can also enhance the edges of structures in an image. Edge-enhancement increases the statistical noise in the image. Spatial filtering is discussed in detail in Chapter 11.

A computer can calculate physiologic performance indices from image data. For example, the estimation of the left ventricular ejection fraction from nuclear gated cardiac blood pool image sequences is described in Chapter 21. In some cases, these data may be displayed graphically, such as the time-activity curves of a bolus of a radiopharmaceutical passing through the kidneys.

In x-ray and nuclear CT, sets of projection images are acquired from different angles about the long axes of patients. From these sets of projection images, computers calculate cross-sectional images depicting tissue linear attenuation coefficients (x-ray CT) or radionuclide concentration (SPECT and PET). The methods

by which volumetric data sets are reconstructed from projection images are discussed in detail in Chapters 13 and 22.

From the volumetric data sets produced by tomography, pseudo-three-dimensional images can be obtained, by methods called surface and volume rendering. These are discussed further in Chapter 13.

In *image co-registration*, image data from one medical imaging modality is superimposed on a spatially congruent image from another modality. Image co-registration is especially useful when one of the modalities, such as SPECT or PET, provides physiologic information lacking anatomic detail and the other modality, often MRI or x-ray CT, provides inferior physiologic information, but superb depiction of anatomy. When the images are obtained using two separate devices, co-registration poses formidable difficulties in matching the position, orientation, and magnification of one image to the other. To greatly simplify co-registration, PET and SPECT devices have been developed that incorporate x-ray CT machines.

Computer-Aided Detection

Computer-aided detection, also known as computer-aided diagnosis, uses a computer program to detect features likely to be of clinical significance in images. Its purpose is not to replace the interpreting physician, but to assist by calling attention to structures that might have been overlooked. For example, software is available to automatically locate structures suggestive of masses, clusters of microcalcifications, and architectural distortions in mammographic images. Computer-aided detection can improve the sensitivity of the interpretation, but also may reduce the specificity. (Sensitivity and specificity are defined in Chapter 10, section 10.7.)

Display

Conversion of a Digital Image into an Analog Video Signal

A computer's video interface converts a digital image into an analog video signal that is displayed by a video monitor. The video interface contains one or more digital-to-analog converters (DACs). Most video interfaces also contain memories to store the image or images being displayed. For a digital image to be displayed, it is first sent to this display memory by the CPU. Once it is stored in this memory, the video interface of the computer reads the pixels in the image sequentially, in the raster pattern shown in Fig. 4-11, and sends these digital numbers to the DAC. The

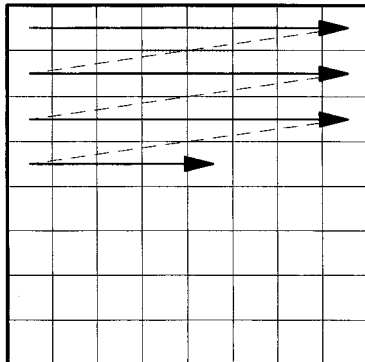


FIGURE 4-11. Order of conversion of pixels in a digital image to analog voltages to form a video signal (raster scan).

DAC converts each digital number to an analog voltage. Other circuits in the video interface add synchronization pulses to this analog signal, producing an analog video signal that may be displayed on a video monitor. The video interface may contain additional circuits that enable the user to adjust the contrast of the image selected for display.

Gray-Scale Cathode Ray Tube Monitors

Radiographic images, consisting of rectangular arrays of numbers, do not inherently have the property of color and thus the images from most modalities are usually displayed on gray-scale monitors. As will be discussed later, gray-scale monitors provide greater ranges of brightness and thus greater dynamic ranges than do color monitors. However, some modalities, in particular nuclear medicine and ultrasound, use color monitors to provide false color displays (described below). A video monitor may use a cathode ray tube or a flat-panel display to create a visible image.

A cathode ray tube (CRT) is an evacuated container, usually made of glass, with components that generate, focus, and modulate the intensity of one or more electron beams directed onto a fluorescent screen. The CRT of a gray-scale monitor (Fig. 4-12) generates a single electron beam. When a potential difference (voltage) is applied between a pair of electrodes, the negative electrode is called the *cathode* and the positive electrode is called the *anode*. If electrons are released from the cathode, the electric field accelerates them toward the anode. These electrons are called *cathode rays*. The source of the electrons, in most CRTs, is a cathode that is heated by an electrical current, and the fluorescent screen is the anode. A large potential difference, typically 10 to 30 kV, applied between these two electrodes creates an electric field that accelerates the electrons to high kinetic energies and directs them onto the screen. Between the cathode and the screen is a grid electrode. A voltage applied to the grid electrode is used to vary the electric current (electrons per sec-

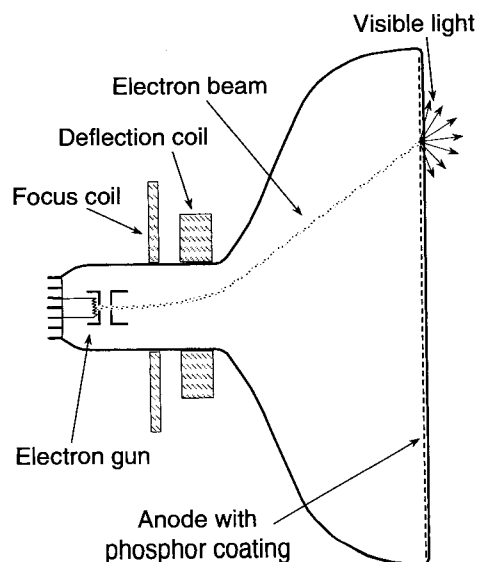


FIGURE 4-12. Gray-scale CRT video monitor. The electron gun contains a heated cathode, the source of the electrons in the beam, and a grid electrode, which modulates the intensity (electrical current) of the beam. This CRT uses an electromagnet to focus the beam and two pairs of electromagnets to steer the beam. (One pair, not shown, is perpendicular to the pair in the diagram.) Alternatively, electrodes placed inside the CRT can perform these functions.

ond) of the beam. The electron beam is focused by either focusing electrodes or an electromagnet into a very narrow cross section where it strikes the screen. The electrons deposit their energy in the phosphor of the screen, causing the emission of visible light. The intensity of the light is proportional to the electric current in the beam, which is determined by the analog video voltage signal applied to the grid electrode. The electron beam can be steered in two perpendicular directions, either by two pairs of electrodes or by two pairs of electromagnets. In a CRT video monitor, the beam is steered in a raster pattern, as shown in Fig. 4-11.

Color CRT Monitors

The CRT of a color monitor is similar to that of a gray-scale monitor, but has three electron guns instead of one, producing three electron beams. The three electron beams are very close to one another and are steered as one, but their intensities are modulated separately. Each pixel of the fluorescent screen contains three distinct regions of phosphors, emitting red, green, and blue light. Just before the screen is a thin metal sheet, called a shadow mask, containing holes positioned so that each electron beam only strikes the phosphor emitting a single color. Thus one of the electron beams creates red light, another creates green light, and the third produces blue light. A mixture of red, green, and blue light, in the proper proportion, can create the perception of any color by the human eye.

Flat-Panel Monitors

Most flat-panel monitors today use liquid crystal displays (LCDs). The liquid crystal material of an LCD does not produce light. Instead, it modulates the intensity of light from another source. The brightest LCDs are backlit. A backlit LCD consists of a uniform light source, typically containing fluorescent tubes and a layer of diffusing material; two polarizing filters; and a layer of liquid crystal material between the two polarizing filters (Fig. 4-13).

Visible light, like any electromagnetic radiation, may be considered to consist of oscillating electrical and magnetic fields, as described in Chapter 2, section 2.1. Unpolarized light consists of light waves whose oscillations are randomly oriented. A polarizing filter is a layer of material that permits components of light waves oscillating in one direction to pass, but absorbs components oscillating in the perpendicular direction. When unpolarized light is incident on a single layer of polarizing material, the intensity of the transmitted light is half that of the incident light. When a second polarizing filter is placed in the beam, parallel to the first filter, the intensity of the beam transmitted through the second filter depends on the orientation of the second filter with respect to the first. If the both filters have the same orientation, the intensity of the beam transmitted through both filters is almost the same as that transmitted through the first. As the second filter is rotated with respect to the first, the intensity of the transmitted light is reduced. When the second filter is oriented so that its polarization is perpendicular to that of the first filter, no light is transmitted.

In an LCD display, the light transmitted through the first filter is polarized. Next, the light passes through a thin layer of liquid crystal material contained between two glass plates. The liquid crystal material consists of rod-shaped molecules aligned so they are parallel. When a voltage is applied across a portion of the

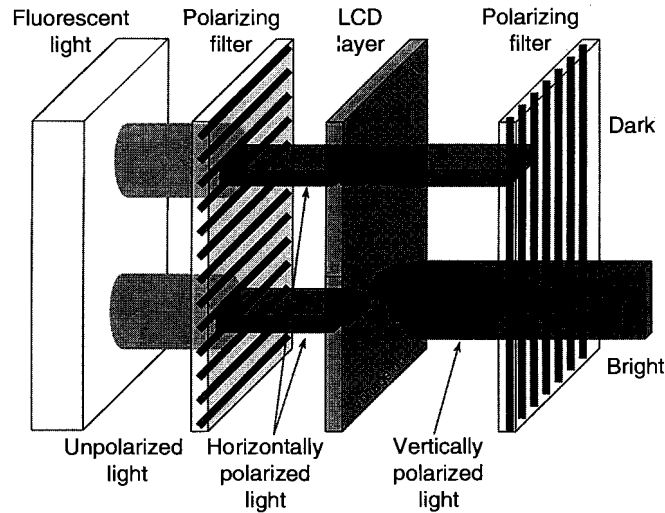


FIGURE 4-13. Two pixels of a gray-scale backlit liquid crystal display video monitor. On each side of the liquid crystal (LC) layer in each pixel is a transparent electrode. In the top pixel, no voltage is applied across the LC layer. The polarization of the light is unchanged as it passes through the LC layer, causing most of the light to be absorbed by the second polarizing filter and thereby producing a dark pixel. In the bottom pixel, a voltage is applied across the LC layer. The applied voltage causes the molecules in the LC layer to twist. The twisted molecules change the polarization of the light, enabling it to pass through the second filter with little reduction in intensity and causing the pixel to appear bright. (The bars shown in the polarizing filters are merely an artist's device to indicate the direction of the polarization, as is depicting the polarized light as a ribbon.)

liquid crystal layer, the molecules twist, changing the polarization of the light by an angle equal to the amount of twist. The light then passes through the second polarizing filter. This filter can be oriented so that pixels are bright when no voltage is applied, or so that pixels are black when no voltage is applied. Figure 4-13 shows the filters oriented so that pixels are dark when no voltage is applied. Thus, when no voltage is applied to a pixel, the light is polarized by the first filter, passes through the liquid crystal material, and then absorbed by the second polarizing filter, resulting in a dark pixel. As the voltage applied to the pixel is increased, the liquid crystal molecules twist, changing the polarization of the light and decreasing the fraction absorbed by the second filter, thereby making the pixel brighter.

Color LCD monitors have an additional layer containing color filters. Each pixel consists of three adjacent elements, one containing a filter transmitting only red light, the second transmitting green light, and the third transmitting blue light. As mentioned previously, a mixture of red, green, and blue light can create the perception of any color.

In theory, each pixel of a flat-panel display could be controlled by its own electrical pathway. If this were done, a gray-scale display of a thousand by a thousand pixels would contain a million electrical pathways and a color display of the same pixel format would have three times as many. In practice, flat-panel displays are matrix controlled, with one conduction pathway serving each row of pixels and one serving each column. For a thousand by a thousand pixel gray-scale display, only a thousand row pathways and a thousand column pathways are required. A signal is sent to a specific pixel by simultaneously providing voltages to the row conductor and the column conductor for that pixel.

The intensity of each pixel must be maintained while signals are sent to other pixels. In active matrix LCDs, each pixel is provided a transistor and capacitor (or three transistors and capacitors for a color display). The electrical charge stored on the capacitor maintains the voltage signal for the pixel while signals are sent to other pixels. The transistors and capacitors are constructed on a sheet of glass or quartz coated with silicon. This sheet is incorporated as a layer within the LCD. Active matrix LCDs are also called thin film transistor (TFT) displays. TFT technology, without the polarizing filters and liquid crystal material, is used in flat-panel x-ray image receptors and is discussed in Chapter 11. Flat-panel monitors are lighter, require less space, consume less electrical power, and generate less heat than CRT monitors. A disadvantage of LCD displays is that the image can be viewed only from a narrow range of angles, whereas CRT displays can be viewed over a very wide range. Another disadvantage of LCD displays is inoperative pixels. Nearly all LCD monitors have some nonfunctional pixels. These may be always on or always off. The performance characteristics of video monitors are described in Chapter 17.

Contrast Enhancement

To display a digital image, pixel values from the minimal pixel value to the maximal value in the image must be displayed. For displaying the image, a range of video intensities from the darkest to the brightest the video monitor can produce is available. There is a great amount of choice in how the mapping from pixel value to video intensity is to be performed. Selecting a mapping that optimizes the display of important features in the image is called *contrast enhancement*.

There are two methods for contrast enhancement available on most medical image-processing computers: translation table selection and windowing. Both of these may be considered to be methods image processing. However, they are intimately connected with displaying images and so are discussed here. Furthermore, both methods may be applied to the entire image in display memory before the pixel values are sent, one at a time, to form a video signal, or they may be performed on each pixel value, one at a time, after it is selected. The second approach will be described below.

A *translation table* is a lookup table that the video interface uses to modify each pixel value before it is sent to the DAC and becomes an analog signal. Figure 4-14

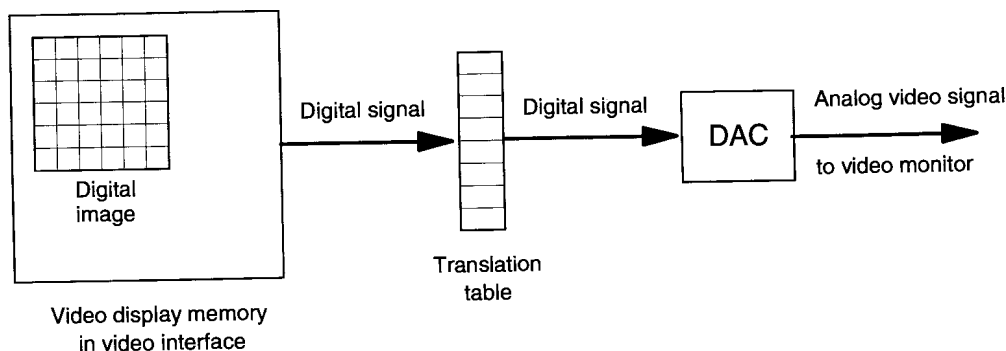


FIGURE 4-14. Video interface showing function of a translation table. An image is transferred to the memory of the video interface for display. The video interface selects pixels in a raster pattern, and sends the pixel values, one at a time, to the lookup table. The lookup table produces a digital value indicating video intensity to the DAC. The DAC converts the video intensity from digital to analog form.

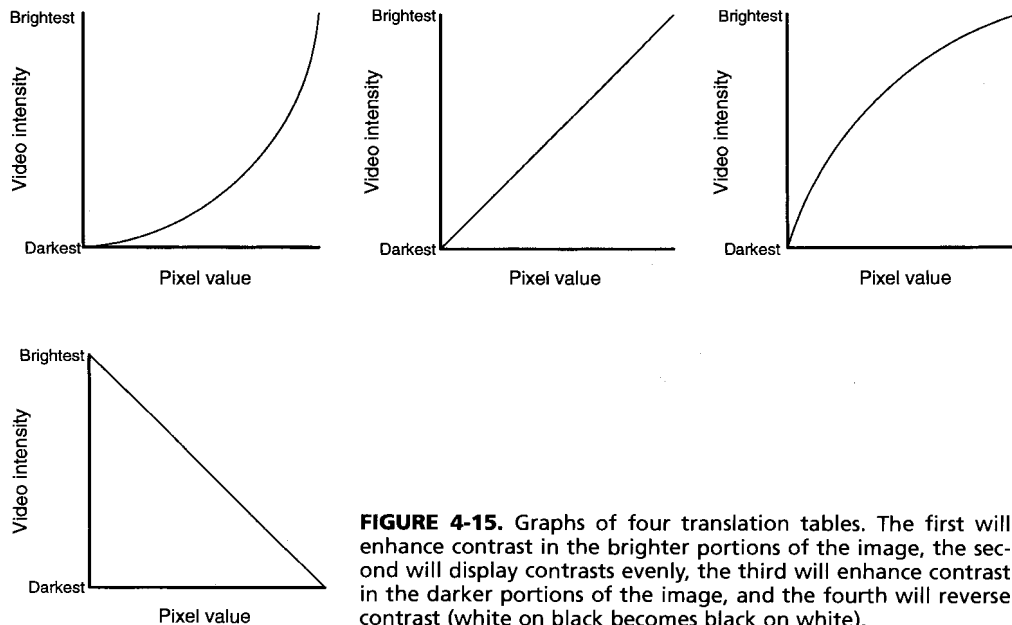


FIGURE 4-15. Graphs of four translation tables. The first will enhance contrast in the brighter portions of the image, the second will display contrasts evenly, the third will enhance contrast in the darker portions of the image, and the fourth will reverse contrast (white on black becomes black on white).

illustrates the use of a translation table. When an image is displayed, each pixel value is sent to the translation table. For every possible pixel value, the table contains a number representing a video intensity. When the translation table receives a pixel value, it looks up the corresponding video intensity and sends this digital number to the DAC. The DAC then converts this number to an analog voltage signal. Figure 4-15 shows four translation tables in graphical form.

Windowing is the second method for contrast enhancement. The viewer may choose to use the entire range of video intensities to display just a portion of the total range of pixel values. For example, an image may have a minimal pixel value of 0 and a maximum of 200. If the user wishes to enhance contrast differences in the brighter portions of the image, a window from 100 to 200 might be selected. Then, pixels with values from 0 to 100 are displayed at the darkest video intensity and so will not be visible on the monitor. Pixels with values of 200 will be displayed at the maximal brightness, and the pixels with values between 100 and 200 will be assigned intensities determined by the current translation table. Figure 4-16 illustrates windowing.

CT and MRI scanners and digital radiography machines usually have level and window controls. The level control determines the midpoint of the pixel values to be displayed and the window control determines the range of pixel values about the level to be displayed. Some nuclear medicine computers require the user to select the lower level (pixel value below which all pixels are displayed at the darkest video intensity) and the upper level (pixel value above which all pixel values are set to the brightest video intensity):

$$\text{Lower level} = \text{Level} - \text{Window}/2$$

$$\text{Upper level} = \text{Level} + \text{Window}/2.$$

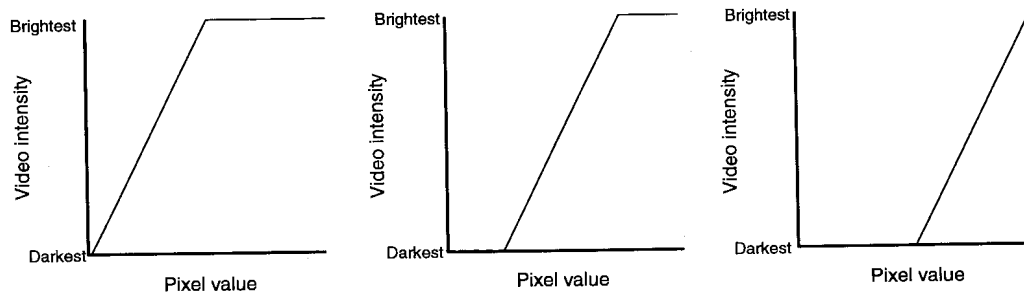


FIGURE 4-16. Windowing an image with pixel values from 0 to 200 to enhance contrast in pixels from 0 to 100 (left), 50 to 150 (center), and 100 to 200 (right).

False Color Displays

As mentioned above, radiographic images do not inherently have the property of color. When color is used to display them, they are called *false color images*. Figure 4-17 shows how false color images are produced. Common uses of false color displays include nuclear medicine, in which color is often used to enhance the perception of contrast, and ultrasound, in which color is used to superimpose flow information on images displaying anatomic information. False color displays are also commonly used in image co-registration, in which an image obtained from one modality is superimposed on a spatially congruent image from another modality.

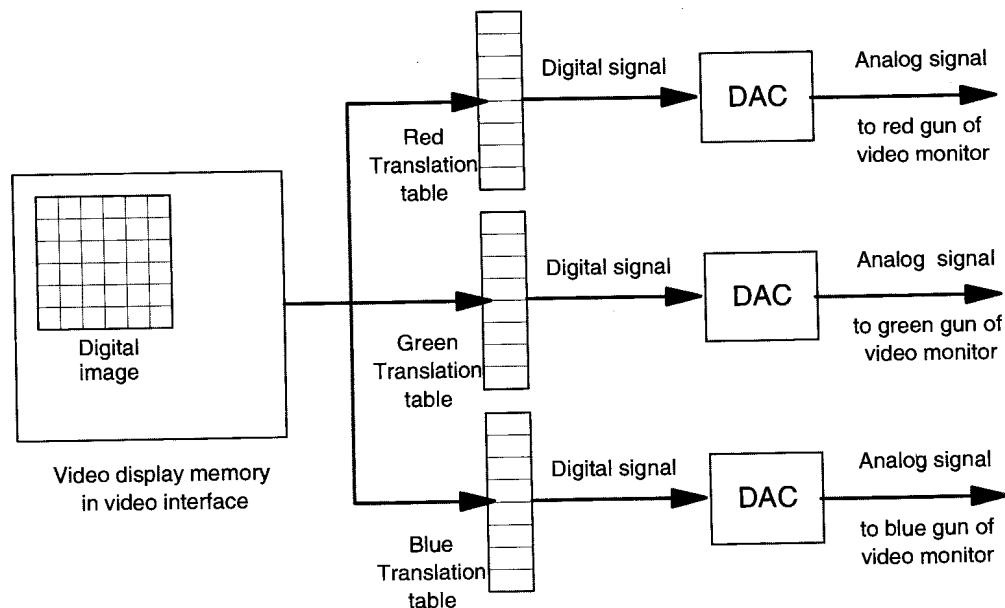


FIGURE 4-17. Creation of a false color display. Each individual pixel value in the image being displayed is used to look up a red, a green, and a blue intensity value. These are simultaneously displayed adjacent to one another within a single pixel of a color video monitor. The mix of these three colors creates the perception of a single color for that pixel.

For example, it is often useful to superimpose nuclear medicine images, which display physiology, but may lack anatomic landmarks, on CT or MRI images. In this case, the nuclear medicine image is usually superimposed in color on a gray-scale MRI or CT image.

Hard-Copy Devices

Digital images may be recorded on photographic film using either a video imager or a laser imager. Most video imagers use the analog video signal produced by the video interface. A video imager consists of a video monitor; a film holder; and one or more lenses, each with a shutter, to focus the video image onto the film, all encased in a light-tight enclosure. A laser imager scans a moving film with a laser beam to produce the image, one row at a time. Most video and laser imagers allow the user to record several images on a single sheet of film. Video and laser imagers are discussed in more detail in Chapter 17.

SECTION
II

DIAGNOSTIC RADIOLOGY

X-RAY PRODUCTION, X-RAY TUBES, AND GENERATORS

X-rays are produced when highly energetic electrons interact with matter and convert their kinetic energy into electromagnetic radiation. A device that accomplishes such a task consists of an electron source, an evacuated path for electron acceleration, a target electrode, and an external energy source to accelerate the electrons. Specifically, the *x-ray tube insert* contains the electron source and target within an evacuated glass or metal envelope; the *tube housing* provides shielding and a coolant oil bath for the tube insert; *collimators* define the x-ray field; and the *generator* is the energy source that supplies the voltage to accelerate the electrons. The generator also permits control of the x-ray output through the selection of voltage, current, and exposure time. These components work in concert to create a beam of x-ray photons of well-defined intensity, penetrability, and spatial distribution. In this chapter, the important aspects of the x-ray creation process, characteristics of the x-ray beam, and details of the equipment are discussed.

5.1 PRODUCTION OF X-RAYS

Bremsstrahlung Spectrum

The conversion of electron kinetic energy into electromagnetic radiation produces x-rays. A simplified diagram of an x-ray tube (Fig. 5-1) illustrates the minimum components. A large voltage is applied between two electrodes (the cathode and the anode) in an evacuated envelope. The cathode is negatively charged and is the *source* of electrons; the anode is positively charged and is the *target* of electrons. As electrons from the cathode travel to the anode, they are accelerated by the electrical potential difference between these electrodes and attain kinetic energy. The electric potential difference, also called the *voltage*, is defined in Appendix A and the SI unit for electric potential difference is the volt (V). The kinetic energy gained by an electron is proportional to the potential difference between the cathode and the anode. For example, the energies of electrons accelerated by potential differences of 20 and 100 kilovolt peak (kVp) are 20 and 100 keV, respectively.

On impact with the target, the kinetic energy of the electrons is converted to other forms of energy. The vast majority of interactions produce unwanted heat by small collisional energy exchanges with electrons in the target. This intense heating limits the number of x-ray photons that can be produced in a given time without

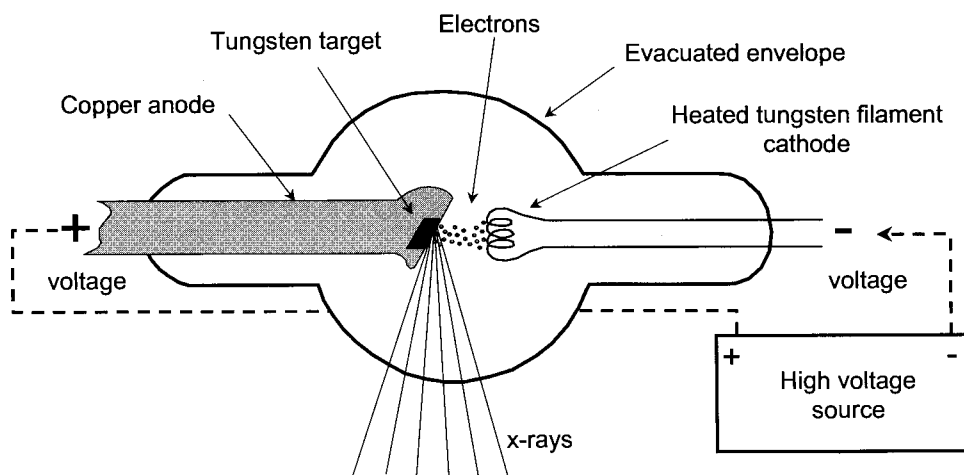


FIGURE 5-1. Minimum requirements for x-ray production include a source and target of electrons, an evacuated envelope, and connection of the electrodes to a high-voltage source.

destroying the target. Occasionally (about 0.5% of the time), an electron comes within the proximity of a positively charged nucleus in the target electrode. Coulombic forces attract and decelerate the electron, causing a significant loss of kinetic energy and a change in the electron's trajectory. An x-ray photon with energy equal to the kinetic energy lost by the electron is produced (conservation of energy). This radiation is termed *bremsstrahlung*, a German word meaning "braking radiation."

The subatomic distance between the bombarding electron and the nucleus determines the energy lost by each electron during the *bremsstrahlung* process, because the coulombic force of attraction increases with the inverse square of the interaction distance. At relatively "large" distances from the nucleus, the coulombic attraction force is weak; these encounters produce low x-ray energies (Fig. 5-2, electron no. 3). For closer interaction distances, the force acting on the electron increases, causing a more dramatic change in the electron's trajectory and a larger loss of energy; these encounters produce higher x-ray energies (see Fig. 5-2, electron no. 2). A direct impact of an electron with the target nucleus results in loss of all of the electron's kinetic energy (see Fig. 5-2, electron no. 1). In this rare situation, the highest x-ray energy is produced.

The probability of an electron's directly impacting a nucleus is extremely low, simply because, at the atomic scale, the atom comprises mainly empty "space" and the nuclear cross-section is very small. Therefore, lower x-ray energies are generated in greater abundance, and the number of higher-energy x-rays decreases approximately linearly with energy up to the maximum energy of the incident electrons. A *bremsstrahlung spectrum* depicts the distribution of x-ray photons as a function of energy. The *unfiltered bremsstrahlung spectrum* (Fig. 5-3a) shows a ramp-shaped relationship between the number and the energy of the x-rays produced, with the highest x-ray energy determined by the peak voltage (kVp) applied across the x-ray tube. Filtration refers to the removal of x-rays as the beam passes through a layer of material. A typical *filtered bremsstrahlung spectrum* (see Fig. 5-3b) shows a distribution with no x-rays below about 10 keV. With filtration, the lower-energy x-rays are preferentially absorbed, and the average x-ray energy is typically about one third to one half of the highest x-ray energy in the spectrum.

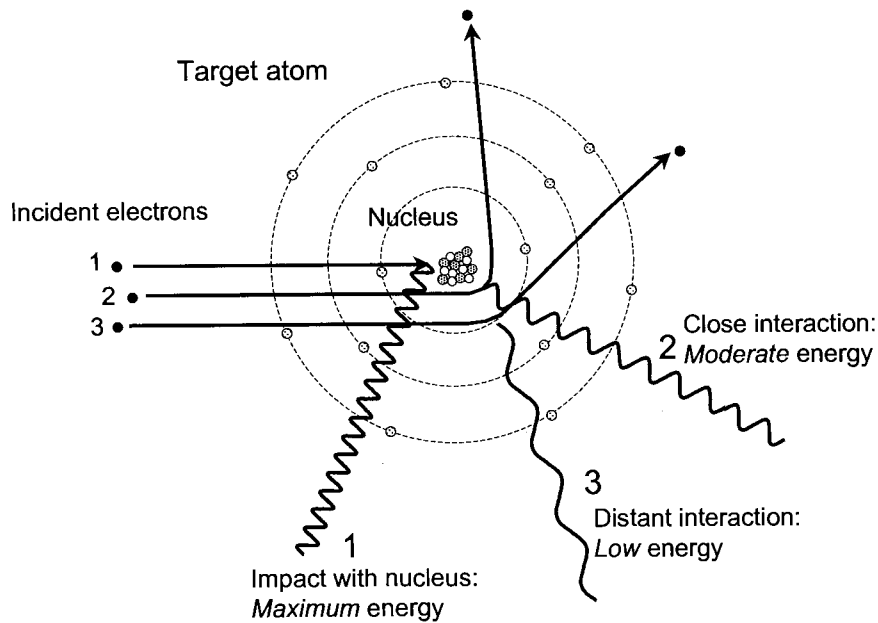


FIGURE 5-2. Bremsstrahlung radiation arises from energetic electron interactions with an atomic nucleus of the target material. In a “close” approach, the positive nucleus attracts the negative electron, causing deceleration and redirection, resulting in a loss of kinetic energy that is converted to an x-ray. The x-ray energy depends on the interaction distance between the electron and the nucleus; it decreases as the distance increases.

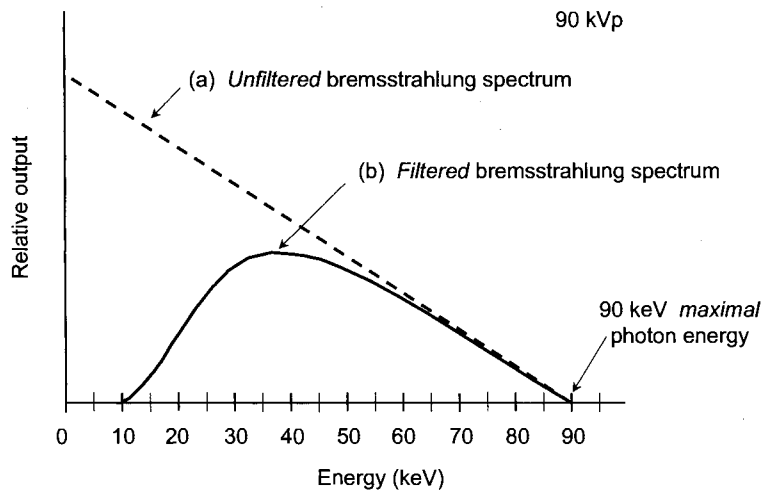


FIGURE 5-3. The bremsstrahlung energy distribution for a 90 kVp acceleration potential. (a) The unfiltered bremsstrahlung spectrum (*dashed line*) shows a greater probability of low-energy x-ray photon production that is inversely linear with energy up to the maximum energy of 90 keV. (b) The filtered spectrum shows the preferential attenuation of the lowest-energy x-ray photons.

Major factors that affect x-ray production efficiency include the atomic number of the target material and the kinetic energy of the incident electrons (which is determined by the accelerating potential difference). The approximate ratio of radiative energy loss caused by bremsstrahlung production to collisional (excitation and ionization) energy loss is expressed as follows:

$$\frac{\text{Radiative energy loss}}{\text{Collisional energy loss}} \cong \frac{E_K Z}{820,000} \quad [5-1]$$

where E_K is the kinetic energy of the incident electrons in keV, and Z is the atomic number of the target electrode material. For 100-keV electrons impinging on tungsten ($Z = 74$), the approximate ratio of radiative to collisional losses is $(100 \times 74)/820,000 \cong 0.009 \cong 0.9\%$; therefore, more than 99% of the incident energy creates heat. However, as the energy of the incident electrons is increased (e.g., to radiation therapy energy levels), x-ray production efficiency increases substantially. For 6-MeV electrons (typical for radiation therapy energies), the ratio of radiative to collisional losses is $(6,000 \times 74)/820,000 \cong 0.54 \cong 54\%$, so excessive heat becomes less of a problem at higher energies.

Characteristic X-Ray Spectrum

Each electron in the target atom has a binding energy that depends on the shell in which it resides. Closest to the nucleus are two electrons in the K shell, which has the highest binding energy. The L shell, with eight electrons, has the next highest binding energy, and so forth. Table 5-1 lists common target materials and the corresponding binding energies of their K , L , and M electron shells. When the energy of an electron incident on the target exceeds the binding energy of an electron of a target atom, it is energetically possible for a collisional interaction to eject the electron and ionize the atom. The unfilled shell is energetically unstable, and an outer shell electron with less binding energy will fill the vacancy. As this electron transitions to a lower energy state, the excess energy can be released as a characteristic x-ray photon with an energy equal to the difference between the binding energies of the electron shells (Fig. 5-4). Binding energies are *unique* to a given element, and so are their differences; consequently, the emitted x-rays have discrete energies that are *characteristic* of that element. For tungsten, an L -shell electron filling a K -shell vacancy results in a characteristic x-ray energy:

$$E_{K\text{-shell}} - E_{L\text{-shell}} = 69.5 \text{ keV} - 10.2 \text{ keV} = 59.3 \text{ keV}$$

Many electron transitions can occur from adjacent and nonadjacent shells in the atom, giving rise to several discrete energy peaks superimposed on the continu-

TABLE 5-1. ELECTRON BINDING ENERGIES (keV) OF COMMON X-RAY TUBE TARGET MATERIALS

Electron Shell	Tungsten	Molybdenum	Rhodium
K	69.5	20.0	23.2
L	12.1/11.5/10.2	2.8/2.6/2.5	3.4/3.1/3.0
M	2.8–1.9	0.5–0.4	0.6–0.2

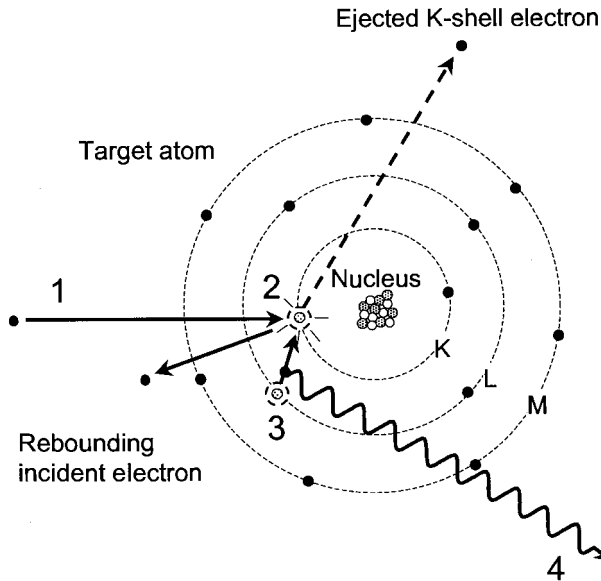


FIGURE 5-4. Generation of a characteristic x-ray in a target atom occurs in the following sequence: (1) The incident electron interacts with the K-shell electron via a repulsive electrical force. (2) The K-shell electron is removed (only if the energy of the incident electron is greater than the K-shell binding energy), leaving a vacancy in the K shell. (3) An electron from the adjacent L shell (or possibly a different shell) fills the vacancy. (4) A K_{α} characteristic x-ray photon is emitted with an energy equal to the difference between the binding energies of the two shells. In this case, a 59.3-keV photon is emitted.

ous bremsstrahlung spectrum. The most prevalent characteristic x-rays in the diagnostic energy range result from K-shell vacancies, which are filled by electrons from the L, M, and N shells. Even unbound electrons outside of the atom have a small probability of filling vacancies. The shell capturing the electron designates the characteristic x-ray transition, and a subscript of α or β indicates whether the transition is from an adjacent shell (α) or nonadjacent shell (β). For example, K_{α} refers to an electron transition from the L to the K shell, and K_{β} refers to an electron transition from M, N, or O shell to K shell. A K_{β} x-ray is more energetic than a K_{α} x-ray. Within each shell (other than the K shell), there are discrete energy subshells, which result in the fine energy splitting of the characteristic x-rays. For tungsten, three prominent lines on the bremsstrahlung spectrum arise from the $K_{\alpha 1}$, $K_{\alpha 2}$, and $K_{\beta 1}$ transitions, as shown in Fig. 5-5. Characteristic x-rays other than those generated

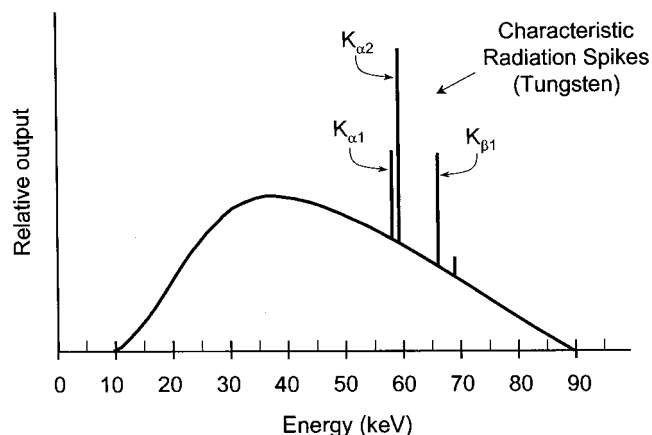


FIGURE 5-5. The filtered spectrum of bremsstrahlung and characteristic radiation from a tungsten target with a potential difference of 90 kVp illustrates specific characteristic radiation energies from K_{α} and K_{β} transitions. Filtration (the preferential removal of low-energy photons as they traverse matter) is discussed in Chapter 3, section 3.

TABLE 5-2. K-SHELL CHARACTERISTIC X-RAY ENERGIES (keV) OF COMMON X-RAY TUBE TARGET MATERIALS^a

Shell Transition	Tungsten	Molybdenum	Rhodium
K _{α1}	59.32	17.48	20.22
K _{α2}	57.98	17.37	20.07
K _{β1}	67.24	19.61	22.72

^aNote: Only prominent transitions are listed.

by *K*-shell transitions are unimportant in diagnostic imaging because they are almost entirely attenuated by the x-ray tube window or added filtration. Table 5-2 lists electron shell binding energies and corresponding *K*-shell characteristic x-ray energies of the common x-ray tube target materials.

Characteristic *K* x-rays are emitted *only* when the electrons impinging on the target *exceed* the binding energy of a *K*-shell electron. Acceleration potentials must be greater than 69.5 kVp for tungsten targets or 20 kVp for molybdenum targets to produce *K* characteristic x-rays. The number of characteristic x-rays relative to bremsstrahlung x-rays increases with bombarding electron energies above the threshold energy for characteristic x-ray production. For example, at 80 kVp approximately 5% of the total x-ray output in a tungsten anode tube is composed of characteristic radiation, whereas at 100 kVp it increases to about 10%. Figure 5-5 shows a spectrum with bremsstrahlung and characteristic radiation.

Characteristic x-ray production in an x-ray tube is mostly the result of electron–electron interactions. However, bremsstrahlung x-ray–electron interactions via the photoelectric effect also contribute to characteristic x-ray production.

5.2 X-RAY TUBES

The x-ray tube provides an environment for x-ray production via bremsstrahlung and characteristic radiation mechanisms. Major components are the *cathode*, *anode*, *rotor/stator*, *glass (or metal) envelope*, and *tube housing* (Fig. 5-6). For diagnostic imaging, electrons from the cathode filament are accelerated toward the anode by a peak voltage ranging from 20,000 to 150,000 V (20 to 150 kVp). The *tube current* is the rate of electron flow from the cathode to the anode, measured in milliamperes (mA), where $1 \text{ mA} = 6.24 \times 10^{15} \text{ electrons/sec}$.

For continuous fluoroscopy the tube current is typically 1 to 5 mA, and for projection radiography tube currents from 100 to 1,000 mA are used with short exposure times (less than 100 msec). The kVp, mA, and exposure time are the three major selectable parameters on the x-ray generator control panel that determine the x-ray beam characteristics (quality and quantity). These parameters are discussed further in the following sections.

Cathode

The source of electrons in the x-ray tube is the cathode, which is a helical filament of tungsten wire surrounded by a *focusing cup* (Fig. 5-7). This structure is electri-

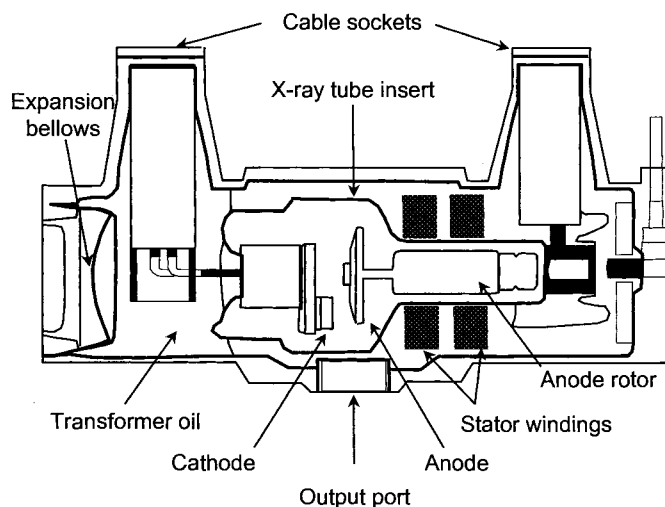


FIGURE 5-6. The major components of a modern x-ray tube and housing assembly.

cally connected to the filament circuit. The filament circuit provides a voltage up to about 10 V to the filament, producing a current up to about 7 A through the filament. Electrical resistance heats the filament and releases electrons via a process called *thermionic emission*. The electrons liberated from the filament flow through the vacuum of the x-ray tube when a positive voltage is placed on the anode with respect to the cathode. Adjustments in the filament current (and thus in the filament temperature) control the tube current. A trace of thorium in the tungsten filament increases the efficiency of electron emission and prolongs filament life.

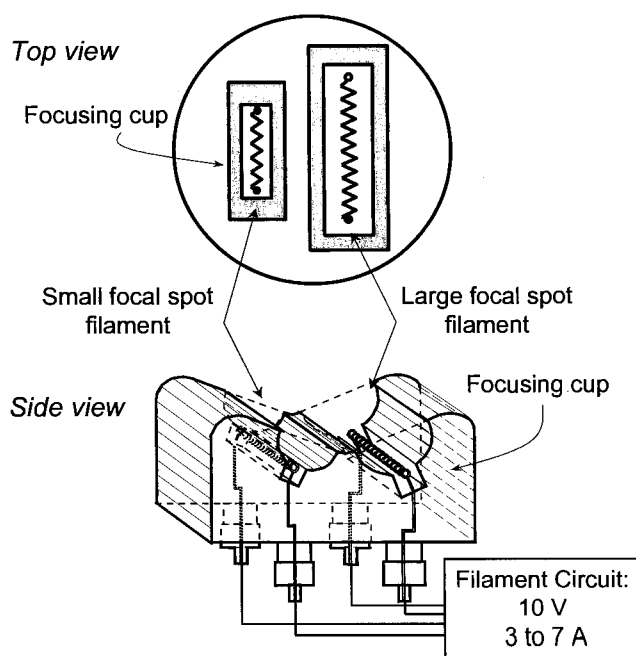


FIGURE 5-7. The x-ray tube cathode structure consists of the filament and the focusing (or cathode) cup. Current from the filament circuit heats the filament, which releases electrons by thermionic emission.

The focusing cup, also called the *cathode block*, surrounds the filament and shapes the electron beam width. The voltage applied to the cathode block is typically the same as that applied to the filament. This shapes the lines of electrical potential to focus the electron beam to produce a small interaction area (focal spot) on the anode. A *biased* x-ray tube uses an insulated focusing cup with a more negative voltage (about 100 V less) than the filament. This creates a tighter electric field around the filament, which reduces spread of the beam and results in a smaller focal spot width (Fig. 5-8). Although the width of the focusing cup slot determines the focal spot width, the filament length determines the focal spot length. X-ray tubes for diagnostic imaging typically have two filaments of different lengths, each in a slot machined into the focusing cup. Selection of one or the other filament determines the area of the electron distribution (small or large focal spot) on the target.

The filament current determines the filament temperature and thus the rate of thermionic electron emission. As the electrical resistance to the filament current heats the filament, electrons are emitted from its surface. When no voltage is applied between the anode and the cathode of the x-ray tube, an electron cloud, also called a *space charge cloud*, builds around the filament. Applying a positive high voltage to the anode with respect to the cathode accelerates the electrons toward the anode and produces a tube current. Small changes in the filament current can produce relatively large changes in the tube current (Fig. 5-9).

The existence of the space charge cloud shields the electric field for tube voltages of 40 kVp and lower, and only a portion of the free electrons are instantaneously accelerated to the anode. When this happens, the operation of the x-ray tube is *space charge limited*, which places an upper limit on the tube current, regardless of the filament current (see the 20- and 40-kVp curves in Fig. 5-9). Above 40 kVp, the space charge cloud effect is overcome by the applied potential difference and the tube current is limited only by the emission of electrons from the filament. Therefore, the filament current controls the tube current in a predictable way (*emission-limited* operation). The tube current is five to ten times *less* than the filament

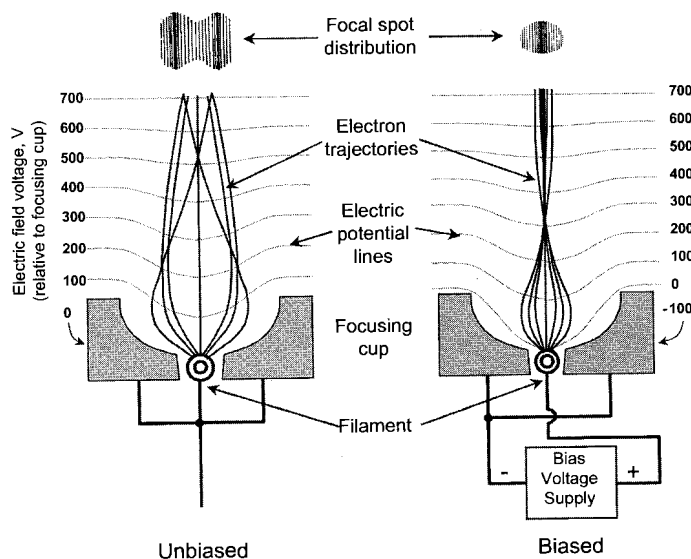


FIGURE 5-8. The focusing cup shapes the electron distribution when it is at the same voltage as the filament (**left**). Isolation of the focusing cup from the filament and application of a negative bias voltage (approximately -100 V) reduces electron distribution further by increasing the repelling electric fields surrounding the filament and modifying the electron trajectories. At the top of the figure, typical electron distributions that are incident on the target anode (the focal spot) are shown for unbiased and biased focusing cups.

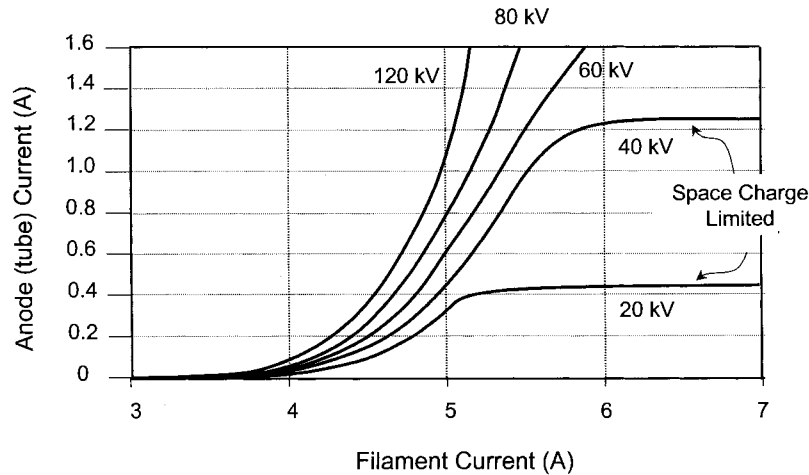


FIGURE 5-9. Relationship of tube current to filament current for various tube voltages shows a dependence of approximately $kVp^{1.5}$. For tube voltages 40 kVp and lower, the space charge cloud shields the electric field so that further increases in filament current do not increase the tube current. This is known as “space charge–limited” operation. Above 40 kVp, the filament current limits the tube current; this is known as “emission–limited” operation.

current in the emission-limited range. Higher kVp produces slightly higher tube current for the same filament current; for example, at 5 A filament current, 80 kVp produces ~800 mA and 120 kVp produces ~1,100 mA, approximately as $kVp^{1.5}$. This relationship does not continue indefinitely. Beyond a certain kVp, saturation occurs whereby all of the emitted electrons are accelerated toward the anode and a further increase in kVp does not significantly increase the tube current.

Anode

The anode is a metal target electrode that is maintained at a positive potential difference relative to the cathode. Electrons striking the anode deposit the most of their energy as heat, with a small fraction emitted as x-rays. Consequently, the production of x-rays, in quantities necessary for acceptable image quality, generates a large amount of heat in the anode. To avoid heat damage to the x-ray tube, the rate of x-ray production must be limited. Tungsten (W, $Z = 74$) is the most widely used anode material because of its high melting point and high atomic number. A tungsten anode can handle substantial heat deposition without cracking or pitting of its surface. An alloy of 10% rhenium and 90% tungsten provides added resistance to surface damage. The high atomic number of tungsten provides better bremsstrahlung production efficiency compared with low- Z elements (see Equation 5-1).

Molybdenum (Mo, $Z = 42$) and rhodium (Rh, $Z = 45$) are used as anode materials in mammographic x-ray tubes. These materials provide useful characteristic x-rays for breast imaging (see Table 5-2). Mammographic tubes are described further in Chapter 8.

Anode Configurations

X-ray tubes have stationary and rotating anode configurations. The simplest type of x-ray tube has a stationary (i.e., fixed) anode. It consists of a tungsten insert embedded in a copper block (Fig. 5-10). The copper serves a dual role: it supports the tungsten target, and it removes heat efficiently from the tungsten target. Unfortunately, the small target area limits the heat dissipation rate and consequently limits the maximum tube current and thus the x-ray flux. Many dental x-ray units, portable x-ray machines, and portable fluoroscopy systems use fixed anode x-ray tubes.

Despite their increased complexity in design and engineering, rotating anodes are used for most diagnostic x-ray applications, mainly because of their greater heat loading and consequent higher x-ray output capabilities. Electrons impart their energy on a continuously rotating target, spreading thermal energy over a large area and mass of the anode disk (Fig. 5-11). A bearing-mounted *rotor* assembly supports the anode disk within the evacuated x-ray tube insert. The rotor consists of copper bars arranged around a cylindrical iron core. A series of electromagnets surrounding the rotor outside the x-ray tube envelope makes up the *stator*, and the combination is known as an *induction motor*. Alternating current passes through the stator windings and produces a “rotating” magnetic field (Fig. 5-12), which induces an electrical current in the rotor’s copper bars (see later discussion). This current induces an opposing magnetic field that pushes the rotor and causes it to spin. Rotation speeds are 3,000 to 3,600 (low speed) or 9,000 to 10,000 (high speed) revolutions per minute (rpm). Low or high speeds are achieved by supplying, respectively, a single-phase (60 Hz) or a three-phase (180 Hz) alternating current to the stator windings. X-ray machines are designed so that the x-ray tube will not be energized if the anode is not up to full speed; this is the cause for the short delay (1 to 2 seconds) when the x-ray tube exposure button is pushed.

Rotor bearings are heat sensitive and are often the cause of x-ray tube failure. Bearings are in the high-vacuum environment of the insert and require special heat-insensitive, nonvolatile lubricants. A molybdenum stem attaches the anode to the rotor/bearing assembly, because molybdenum is a very poor heat conductor and reduces heat transfer from the anode to the bearings. Because it is thermally isolated, the anode must be cooled by radiative emission. Heat energy is emitted from the hot anode as infrared radiation, which transfers heat to the x-ray tube insert and ultimately to the surrounding oil bath.

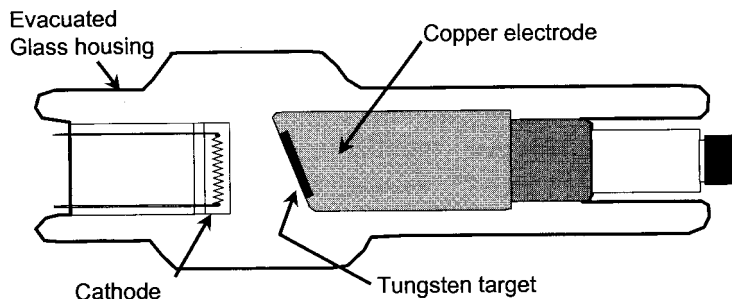


FIGURE 5-10. The anode of a fixed anode x-ray tube consists of a tungsten insert mounted in a copper block. Heat is removed from the tungsten target by conduction into the copper block.

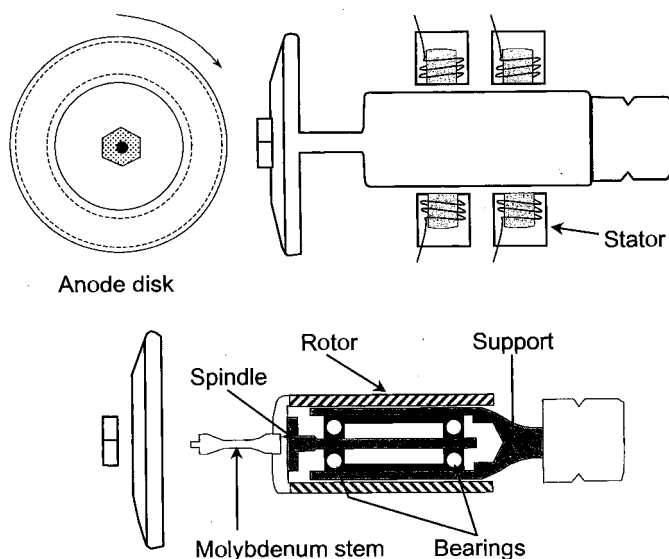


FIGURE 5-11. The anode of a rotating anode x-ray tube comprises a tungsten disk mounted on a bearing-supported rotor assembly (front view, **top left**; side view, **top right**). The rotor consists of a copper and iron laminated core and forms part of an induction motor (the other component is the stator, which exists outside of the insert—see Fig. 5-12). A molybdenum stem (molybdenum is a poor heat conductor) connects the rotor to the anode to reduce heat transfer to the rotor bearings (**bottom**).

The focal track area of the rotating anode is equal to the product of the track length ($2\pi r$) and the track width (Δr), where r is the radial distance from the track to its center. A rotating anode with a 5-cm focal track radius and a 1-mm track width provides a focal track with an annular area 314 times greater than that of a fixed anode with a focal spot area of 1 mm $\times$ 1 mm. The allowable instantaneous heat loading depends on the anode rotation speed and the focal spot area. Faster rotation speeds distribute the heat load over a greater portion of the focal track area.

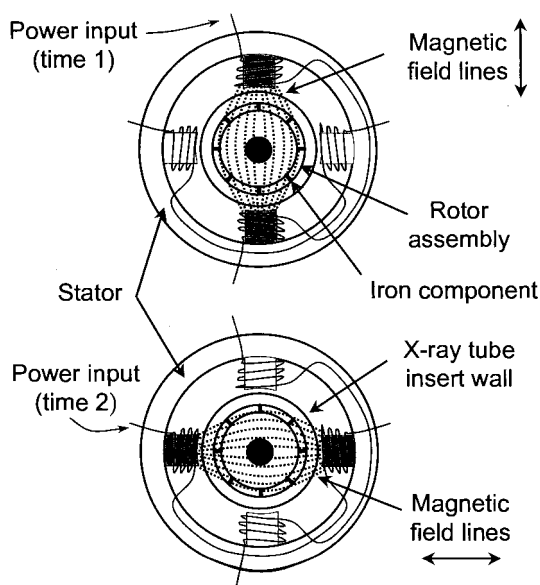


FIGURE 5-12. A cross-sectional end view of the stator/rotor assembly shows electromagnetic coils energized at two different times. Power applied sequentially to pairs of electromagnets produces a rotating magnetic field with an angular frequency of 3,000 (single-phase operation) to 10,000 (three-phase operation) rpm. The magnetic field “drags” the iron components of the rotor to full rotational speed in about 1 second, during which time x-ray exposure is prohibited.

These and other factors determine the maximum heat load possible (see Section 5.8: Power Ratings and Heat Loading).

Anode Angle and Focal Spot Size

The anode angle is defined as the angle of the target surface with respect to the central ray in the x-ray field (Fig. 5-13). Anode angles in diagnostic x-ray tubes, other than some mammography tubes, range from 7 to 20 degrees, with 12- to 15-degree angles being most common. Focal spot size is defined in two ways. The actual focal spot size is the area on the anode that is struck by electrons, and it is primarily determined by the length of the cathode filament and the width of the focusing cup slot. The effective focal spot size is the length and width of the focal spot as projected down the central ray in the x-ray field. The effective focal spot width is equal to the actual focal spot width and therefore is not affected by the anode angle. However, the anode angle causes the effective focal spot length to be smaller than the actual focal spot length. The effective and actual focal spot lengths are related as follows:

$$\text{Effective focal length} = \text{Actual focal length} \times \sin \theta \quad [5-2]$$

where θ is the anode angle. This foreshortening of the focal spot length, as viewed down the central ray, is called the *line focus principle*.

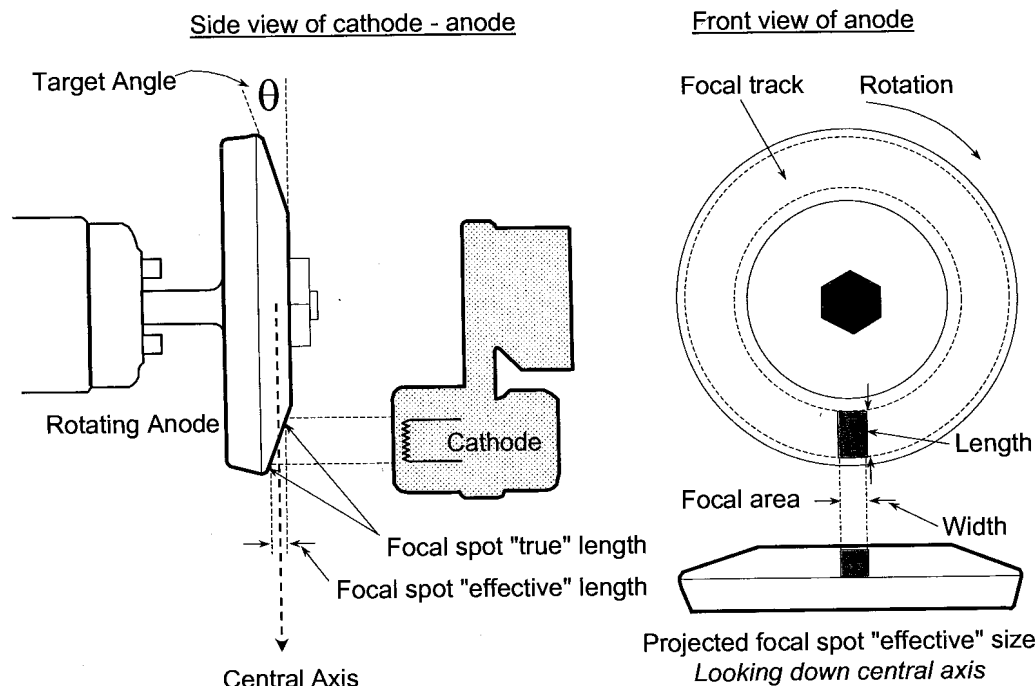


FIGURE 5-13. The anode (target) angle, θ , is defined as the angle of the target surface in relation to the central ray. The focal spot length, as projected down the central axis, is foreshortened, according to the line focus principle (lower right).

Example 1: The actual anode focal area for a 20-degree anode angle is 4 mm (length) by 1.2 mm (width). What is the projected focal spot size at the central axis?

Answer: Effective length = actual length $\times \sin \theta = 4 \text{ mm} \times \sin 20 \text{ degrees} = 4 \text{ mm} \times 0.34 = 1.36 \text{ mm}$; therefore, the projected focal spot size is 1.36 mm (length) by 1.2 mm (width).

Example 2: If the anode angle in Example 1 is reduced to 10 degrees and the actual focal spot size remains the same, what is the projected focal spot size at the central ray?

Answer: Effective length = $4 \text{ mm} \times \sin 10 \text{ degrees} = 4 \text{ mm} \times 0.174 = 0.69 \text{ mm}$; thus, the smaller anode angle results in a projected size of 0.69 mm (length) by 1.2 mm (width) for the same actual target area.

There are three major tradeoffs to consider for the choice of anode angle. A smaller anode angle provides a smaller *effective* focal spot for the same *actual* focal area. (A smaller effective focal spot size provides better spatial resolution, as described in Chapter 6.) However, a small anode angle limits the size of the usable x-ray field owing to cutoff of the beam. Field *coverage* is less for short focus-to-detector distances (Fig. 5-14). The optimal anode angle depends on the clinical imaging application. A small anode angle (approximately 7 to 9 degrees) is desirable for small field-of-view image receptors, such as cineangiographic and neuroangiographic equipment, where field coverage is limited by the image intensifier diameter (e.g., 23 cm). Larger anode angles (approximately 12 to 15 degrees) are necessary for general radiographic work to achieve large field area coverage at short focal spot-to-image distances.

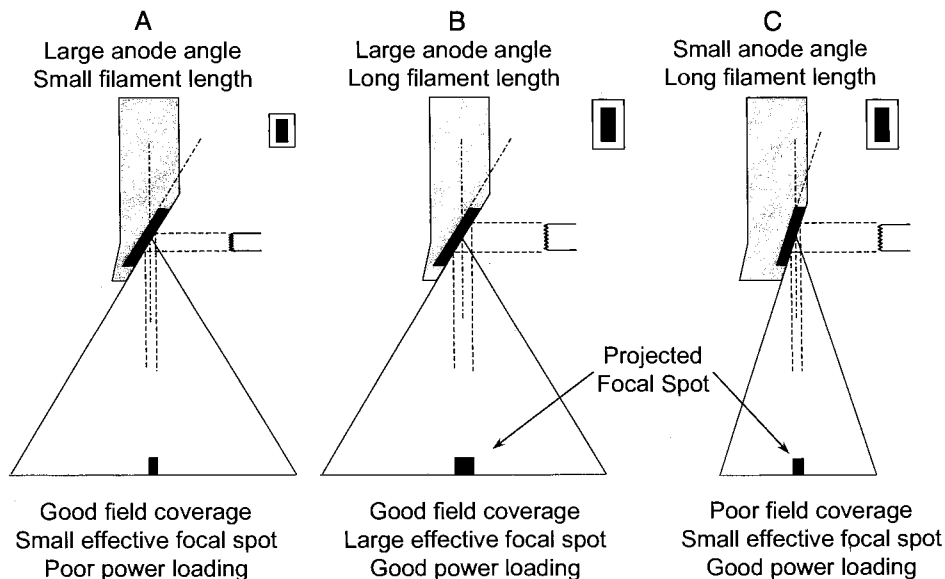


FIGURE 5-14. Field coverage and effective focal spot length vary with the anode angle. **A:** A large anode angle provides good field coverage at a given distance; however, to achieve a small effective focal spot, a small actual focal area limits power loading. **B:** A large anode angle provides good field coverage, and achievement of high power loading requires a large focal area; however, geometric blurring and image degradation occur. **C:** A small anode angle limits field coverage at a given distance; however, a small effective focal spot is achieved with a large focal area for high power loading.

The effective focal spot length varies with the position in the image plane, in the anode-cathode (A-C) direction. Toward the anode side of the field the projected length of the focal spot shortens, whereas it lengthens towards the cathode side of the field (Fig. 5-15). In the width dimension, the focal spot size does not change appreciably with position in the image plane. The nominal focal spot size (width and length) is specified at the central ray of the beam. The central ray is usually a line from the focal spot to the image receptor that is perpendicular to the A-C axis of the x-ray tube and perpendicular to the plane of a properly positioned image receptor. In most radiographic imaging, the central ray bisects the detector field; x-ray mammography is an exception, as explained in Chapter 8.

Tools for measuring focal spot size are the pinhole camera, slit camera, star pattern, and resolution bar pattern (Fig. 5-16A–D). The *pinhole camera* uses a very small circular aperture (10 to 30 μm diameter) in a disk made of a thin, highly attenuating metal such as lead, tungsten, or gold. With the pinhole camera positioned on the central axis between the x-ray source and the detector, an image of the focal spot is recorded. Figure 5-16E illustrates a magnified (2 $\times$) pinhole picture of a focal spot with a typical bi-gaussian (“double banana”) intensity distribution. The *slit camera* (see Fig. 5-16F) consists of a plate made of a highly attenuating metal (usually tungsten) with a thin slit, typically 10 μm wide. In use, the slit camera is positioned above the image receptor, with the center of the slit on the central axis and the slit either parallel or perpendicular to the A-C axis. Measuring the width of the distribution and correcting for magnification (discussed in Chapter 6) yields one dimension of the focal spot. A second radiograph, taken with the slit perpendicular to the first, yields the other dimension of the focal spot. The *star pattern* test tool (see Fig. 5-16G) contains a radial pattern of lead spokes of diminishing width and spacing on a thin plastic disk. Imaging the star pattern at a known magnification and measuring the distance between the outermost blur patterns (areas of unresolved spokes) on the image provides an estimate of the resolving power of the focal spot in the directions perpendicular

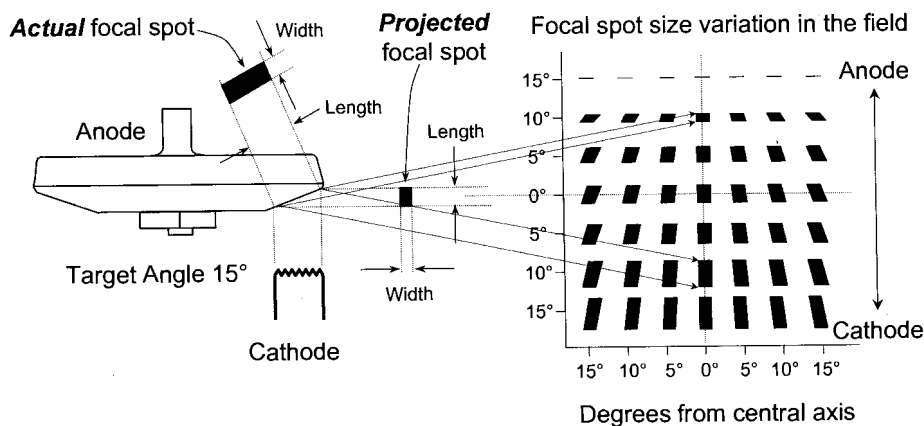


FIGURE 5-15. Variation of the effective focal spot size in the image field occurs along the anode-cathode direction. Focal spot distributions are plotted as a function of projection angle in degrees from the central axis, the parallel (vertical axis), and the perpendicular (horizontal axis).

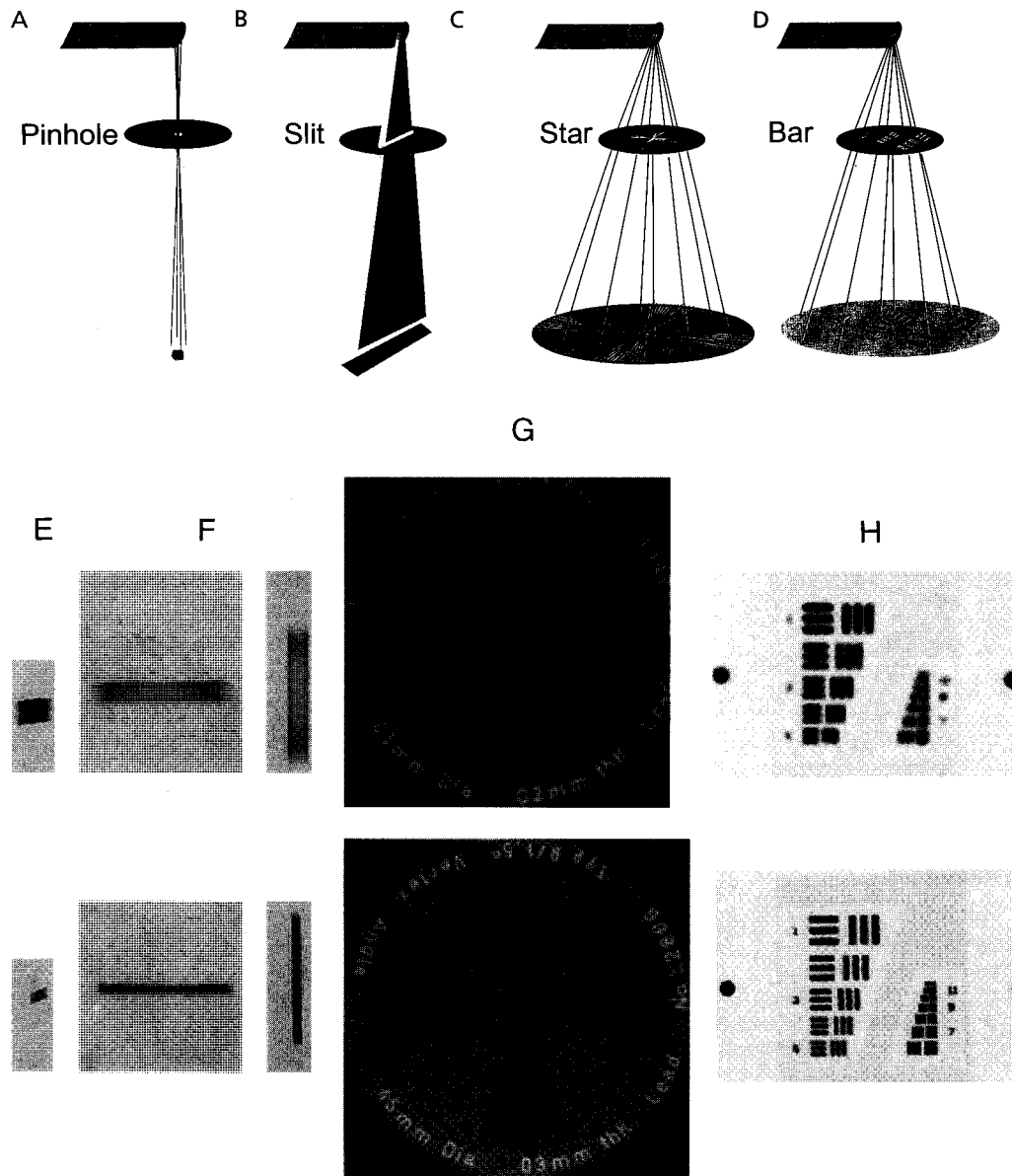


FIGURE 5-16. Various tools allow measurement of the focal spot size, either directly or indirectly. **A** and **E**: Pinhole camera and images. **B** and **F**: Slit camera and images. **C** and **G**: Star pattern and images. **D** and **H**: Resolution bar pattern and images. For **E–H**, the top and bottom images represent the large and small focal spot, respectively.

and parallel to the A-C axis. A large focal spot has a greater blur diameter than a small focal spot. The effective focal spot size can be estimated from the blur pattern diameter and the known magnification. A resolution bar pattern is a simple tool for in-the-field evaluation of focal spot size (Fig. 5-16H). Bar pattern images demonstrate the effective resolution parallel and perpendicular to the A-C axis for a given magnification geometry.

Heel Effect

The *heel effect* refers to a reduction in the x-ray beam intensity toward the anode side of the x-ray field (Fig. 5-17). X-rays are produced isotropically at depth in the anode structure. Photons directed toward the anode side of the field transit a greater thickness of the anode and therefore experience more attenuation than those directed toward the cathode side of the field. For a given field size, the heel effect is less prominent with a longer source-to-image distance (SID), because the image receptor subtends a smaller beam angle. The x-ray tube is best positioned with the cathode over thicker parts of the patient and the anode over the thinner parts, to better balance the transmitted x-ray photons incident on the image receptor.

Off-Focus Radiation

Off-focus radiation results from electrons in the x-ray tube that strike the anode outside the focal spot area. A small fraction of electrons scatter from the target and are accelerated back to the anode outside the focal spot. These electrons create a low-intensity x-ray source over the face of the anode. Off-focus radiation increases patient exposure, geometric blurring, and background fog. A small lead collimator placed near the x-ray tube output port can reduce off-focus radiation by intercepting x-rays that are produced at a large distance from the focal spot. X-ray tubes that have a metal enclosure with the anode at the same electrical potential (i.e., grounded anode) reduce off-focus radiation because stray electrons are just as likely to be attracted to the metal envelope as to the anode.

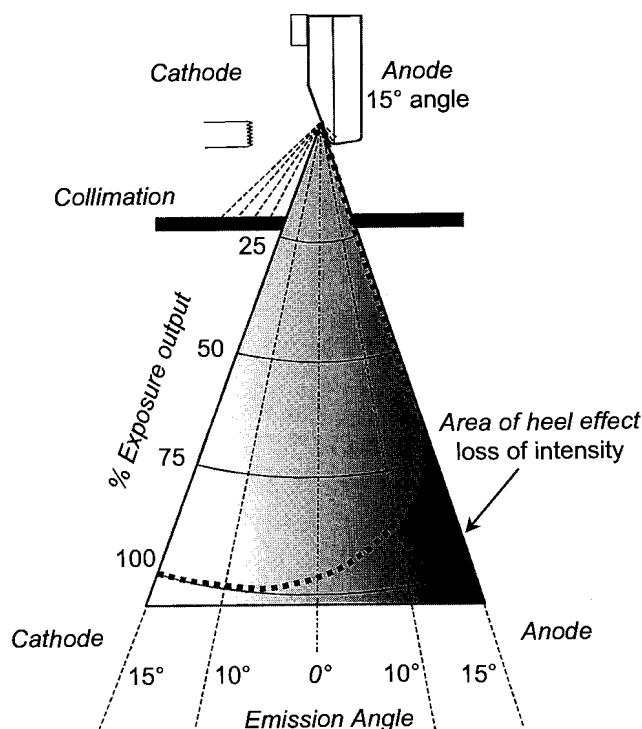


FIGURE 5-17. The heel effect is a loss of intensity on the anode side of the x-ray field of view. It is caused by attenuation of the x-ray beam by the anode.

5.3 X-RAY TUBE INSERT, TUBE HOUSING, FILTRATION, AND COLLIMATION

X-Ray Tube Insert

The *x-ray tube insert* contains the cathode, anode, rotor assembly, and support structures sealed in a glass or metal enclosure (see Fig. 5-6) under a high vacuum. The high vacuum prevents electrons from colliding with gas molecules and is required in all electron beam devices. As x-ray tubes age, trapped gas molecules percolate from tube structures and slightly degrade the vacuum. A “getter” circuit is used to trap gas in the insert.

X-rays are emitted in all directions from the focal spot; however, the x-rays that emerge through the *tube port* constitute the useful beam. Except for mammography x-ray tubes, the port is typically made of the same material as the tube enclosure. Mammography tubes use beryllium ($Z = 4$) in the port to minimize absorption of the low-energy x-rays used in mammography.

X-Ray Tube Housing

The x-ray tube housing supports, insulates, and protects the x-ray tube insert from the environment. The x-ray tube insert is bathed in a special oil, contained within the housing, that provides heat conduction and electrical insulation. A typical x-ray tube housing contains a bellows to allow for oil expansion as it absorbs heat during operation. If the oil heats excessively, the expanding bellows activates a microswitch that shuts off the system until the tube cools.

Lead shielding inside the housing attenuates the x-rays that are emitted in all directions, and of course there is a hole in the shielding at the x-ray tube port. Leakage radiation consists of x-rays that penetrate this lead shielding, and therefore it has a high effective energy. The tube housing must contain sufficient shielding to meet federal regulations that limit the leakage radiation exposure rate to 100 mR/hour at 1 m from the focal spot when the x-ray tube is operated at the maximum kVp (kV_{max} , typically 125 to 150 kVp) and the highest possible continuous current (typically 3 to 5 mA at kV_{max} for most diagnostic tubes).

Special X-Ray Tube Designs

Grid-biased tubes have a focusing cup that is electrically isolated from the cathode filament and maintained at a more negative voltage. When the bias voltage is sufficiently large, the resulting electric field lines shut off the tube current. Turning off the grid bias allows the tube current to flow and x-rays to be produced. Grid biasing requires approximately $-2,000$ V applied to the focusing cup with respect to the filament to switch the x-ray tube current off. The grid-biased tube is used in applications such as pulsed fluoroscopy and cineangiocardiology, where rapid x-ray pulsing is necessary. Biased x-ray tubes are significantly more expensive than conventional, nonbiased tubes.

Mammography tubes are designed to provide the low-energy x-rays necessary to produce optimal mammographic images. As explained in Chapter 8, the main differences between a dedicated mammography tube and a conventional x-ray tube are the target material (molybdenum or rhodium versus tungsten) and the output port (beryllium versus glass or metal insert material).

Filtration

As mentioned earlier, filtration is the removal of x-rays as the beam passes through a layer of material. Attenuation of the x-ray beam occurs because of both the inherent filtration of the tube and added filtration. Inherent filtration includes the thickness (1 to 2 mm) of the glass or metal insert at the x-ray tube port. Glass (SiO_2) and aluminum have similar attenuation properties ($Z = 14$ and $Z = 13$, respectively) and effectively attenuate all x-rays in the spectrum below approximately 15 keV. Dedicated mammography tubes, on the other hand, require beryllium ($Z = 4$) to improve the transmission of low-energy x-rays. Inherent filtration can include attenuation by housing oil and the field light mirror in the collimator assembly.

Added filtration refers to sheets of metal intentionally placed in the beam to change its effective energy. In general diagnostic radiology, added filters attenuate the low-energy x-rays in the spectrum that have virtually no chance of penetrating the patient and reaching the x-ray detector. Because the low-energy x-rays are absorbed by the filters instead of the patient, the radiation dose is reduced. Aluminum (Al) is the most common added filter material. Other common filter materials include copper and plastic (e.g., acrylic). In mammography, molybdenum and rhodium filters are used to help shape the x-ray spectrum (see Chapter 8). Rare earth filters such as erbium are sometimes employed in chest radiography to reduce patient dose.

An indirect measure of the effective energy of the beam is the half value layer (HVL), described in Chapter 3, section 3. The HVL is usually specified in millimeters of aluminum. At a given kVp, added filtration increases the HVL.

In the United States, x-ray system manufacturers must comply with minimum HVL requirements specified in Title 21 of the Code of Federal Regulations (Table 5-3), which have been adopted by many state regulatory agencies. Increases are under consideration (see Table 5-3, right column) that would cause these HVL requirements to conform to standards promulgated by the International Electrotechnical Commission, with the aim of reducing patient radiation dose while maintaining image quality and adequate beam intensity.

Compensation (equalization) filters are used to change the spatial pattern of the x-ray intensity incident on the patient, so as to deliver a more uniform x-ray exposure to the detector. For example, a trough filter used for chest radiography has a centrally located vertical band of reduced thickness and consequently produces greater x-ray fluence in the middle of the field. This filter compensates for the high attenuation of the mediastinum and reduces the exposure latitude incident on the image receptor. Wedge filters are useful for lateral projections in cervical-thoracic spine imaging, where the incident fluence is increased to match the increased tissue thickness encountered (e.g., to provide a low incident flux to the thin neck area and a high incident flux to the thick shoulders). "Bow-tie" filters are used in computed tomography (CT) to reduce dose to the periphery of the patient, where x-ray paths are shorter and fewer x-rays are required. Compensation filters are placed close to the x-ray tube port or just external to the collimator assembly.

Collimators

Collimators adjust the size and shape of the x-ray field emerging from the tube port. The typical collimator assembly is attached to the tube housing at the tube port with a swivel joint. Adjustable parallel-opposed lead shutters define the x-ray field (Fig. 5-18). In the collimator housing, a beam of light reflected by a mirror (of low

TABLE 5-3. MINIMUM HALF VALUE LAYER (HVL) REQUIREMENTS FOR X-RAY SYSTEMS IN THE UNITED STATES (21 CFR 1020.30)^a

X-ray tube voltage (kVp)		Minimum HVL (mm of aluminum)	
Designed Operating Range	Measured Operating Potential	Requirement as of January 2001	Possible Revision
<51	30	0.3	0.3
	40	0.4	0.4
	50	0.5	0.5
51–70	51	1.2	1.3
	60	1.3	1.5
	70	1.5	1.8
>70	71	2.1	2.4
	80	2.3	2.8
	90	2.5	3.2
	100	2.7	3.6
	110	3.0	4.1
	120	3.2	4.5
	130	3.5	5.0
	140	3.8	5.4
	150	4.1	5.9

^aNote: This table does not include values for dental or mammography systems. The values in the column on the right were developed by the staff of the Center for Devices and Radiological Health of the U.S. Food and Drug Administration, as a possible revision to the regulations. These values conform to the standards of the International Electrotechnical Commission.

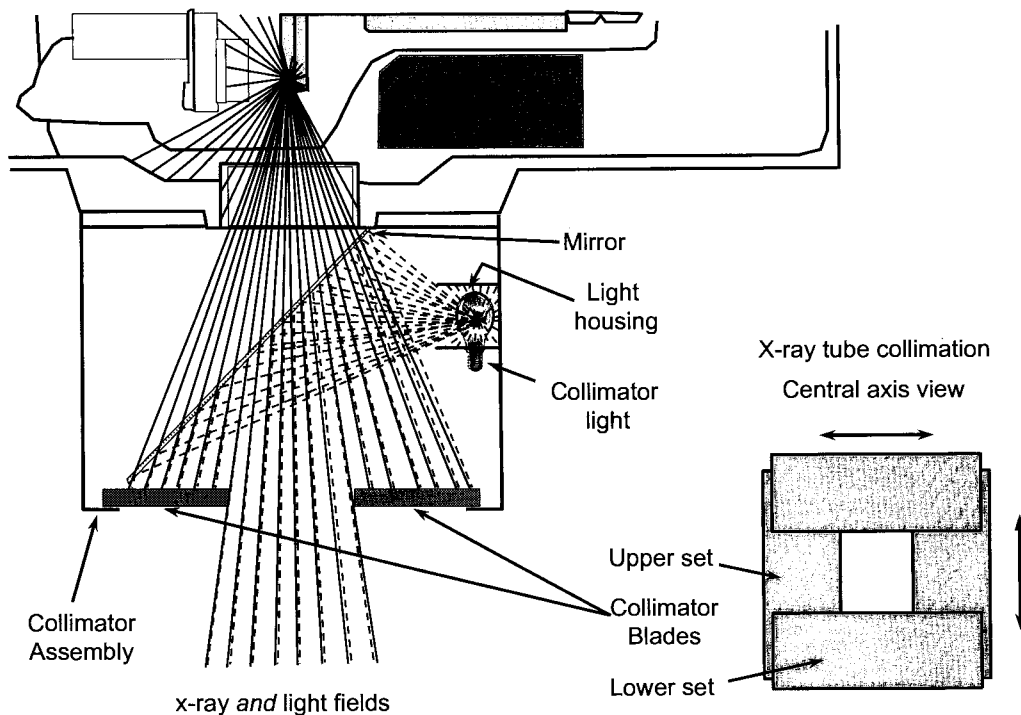


FIGURE 5-18. The x-ray tube collimator assembly is attached to the housing at the tube port, typically on a collar that allows rotational positioning. A light source, positioned at a virtual focal spot location, illuminates the field from a 45-degree angle mirror. Lead collimator blades define both the x-ray and light fields.

x-ray attenuation) mimics the x-ray beam. Thus, the collimation of the x-ray field is identified by the collimator's shadows. Federal regulations (21 CFR 1020.31) require that the light field and x-ray field be aligned so that the sum of the misalignments, along either the length or the width of the field, is within 2% of the SID. For example, at a typical SID of 100 cm (40 inches), the sum of the misalignments between the light field and the x-ray field at the left and right edges must not exceed 2 cm and the sum of the misalignments at the other two edges also must not exceed 2 cm.

Positive beam limitation (PBL) collimators automatically limit the field size to the useful area of the detector. Mechanical sensors in the film cassette holder detect the cassette size and location and automatically adjust the collimator blades so that the x-ray field matches the cassette dimensions. Adjustment to a smaller field area is possible; however, a larger field area requires disabling the PBL circuit.

5.4 X-RAY GENERATOR FUNCTION AND COMPONENTS

Electromagnetic Induction and Voltage Transformation

The principal function of the x-ray generator is to provide current at a high voltage to the x-ray tube. Electrical power available to a hospital or clinic provides up to about 480 V, much lower than the 20,000 to 150,000 V needed for x-ray production. Transformers are principal components of x-ray generators; they convert low voltage into high voltage through a process called *electromagnetic induction*.

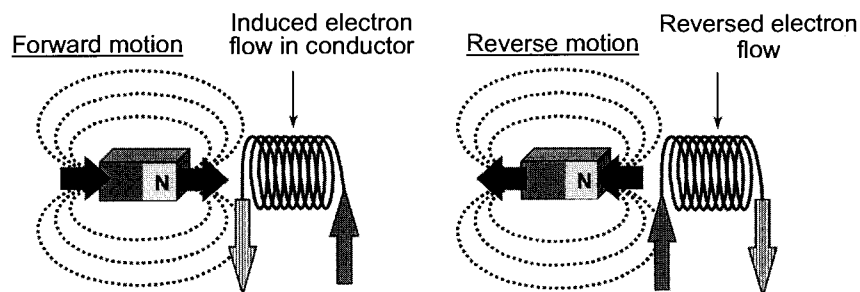
Electromagnetic induction is an effect that occurs with changing magnetic fields and alternating (AC) electrical current. For example, the changing magnetic field from a moving bar magnet induces an electrical potential difference (voltage) and a current in a nearby conducting wire (Fig. 5-19A). As the magnet moves in the opposite direction, away from the wire, the induced current flows in the opposite direction. The magnitude of the induced voltage is proportional to the magnetic field strength.

Electromagnetic induction is reciprocal between electric and magnetic fields. An electrical current (e.g., electrons flowing through a wire) produces a magnetic field, whose magnitude (strength) and polarity (direction) are proportional to the magnitude and direction of the current (see Fig. 5-19B). With an alternating current, such as the standard 60 cycles per second (Hz) AC in North America and 50 Hz AC in most other areas of the world, the induced magnetic field increases and decreases with the current. For "coiled wire" geometry, superimposition of the magnetic fields from adjacent "turns" of the wire increases the amplitude of the overall magnetic field (the magnetic fields penetrate the wire insulation), and therefore the magnetic field strength is proportional to the number of wire turns. Note that for constant-potential direct current (DC), like that produced by a battery, magnetic induction does not occur.

Transformers

Transformers perform the task of "transforming" an alternating input voltage into an alternating output voltage, using the principles of electromagnetic induction. The generic transformer has two distinct, electrically insulated wires wrapped about a common iron core (Fig. 5-20). Input AC power (voltage and current) produces oscillating electrical and magnetic fields. One insulated wire wrapping (the "pri-

A Changing magnetic field induces electron flow:



B Current (electron flow) in a conductor creates a magnetic field;
amplitude and direction determines magnetic field strength and polarity

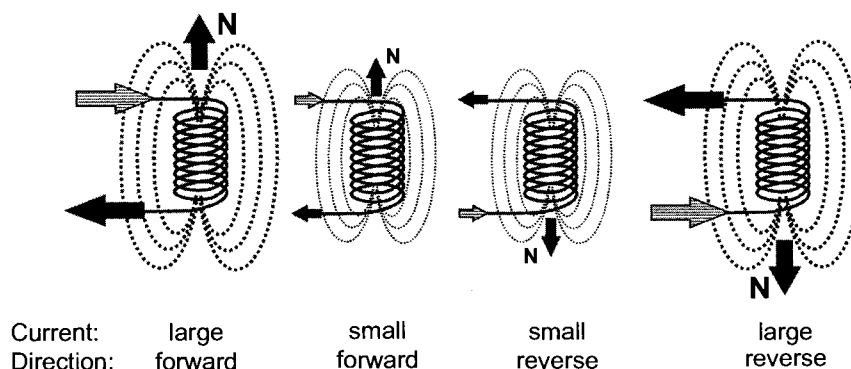


FIGURE 5-19. Principles of electromagnetic induction. **A:** Induction of an electrical current in a wire conductor coil by a moving magnetic field. The direction of the current is dependent on the direction of the magnetic field motion. **B:** Creation of a magnetic field by the current in a conducting coil. The polarity and magnetic field strength are directly dependent on the amplitude and direction of the current.

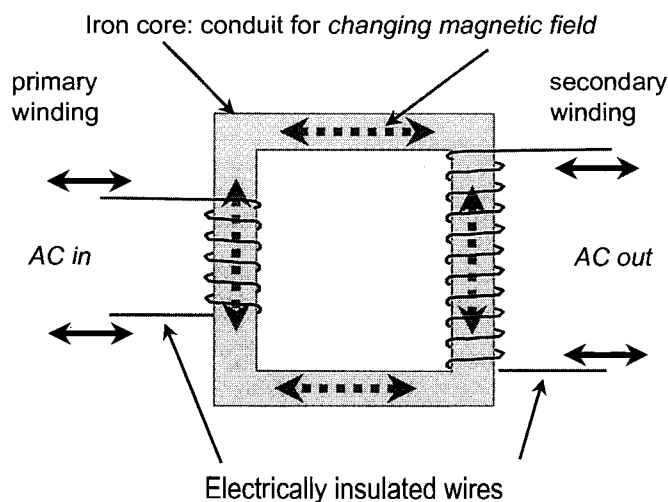


FIGURE 5-20. The basic transformer consists of an iron core, a primary winding circuit, and a secondary winding circuit. An alternating current flowing through the primary winding produces a changing magnetic field. The magnetic field permeates the core and produces an alternating voltage on the secondary winding. This exchange is mediated by the permeability of the magnetic field through wire insulation and in the iron core.

primary winding”) carries the input load (primary voltage and current). The other insulated wire wrapping (“secondary winding”) carries the induced (output) load (secondary voltage and current). The primary and secondary windings are electrically (but not magnetically) isolated by insulated wires. The induced magnetic field strength changes amplitude and direction with the primary AC voltage waveform and freely passes through electrical insulation to permeate the transformer iron core, which serves as a conduit and containment for the oscillating magnetic field. The secondary windings are bathed in the oscillating magnetic field, and an alternating voltage is induced in the secondary windings as a result. The Law of Transformers states that the ratio of the number of coil turns in the primary winding to the number of coil turns in the secondary winding is equal to the ratio of the primary voltage to the secondary voltage:

$$\frac{V_P}{V_S} = \frac{N_P}{N_S} \quad [5-3]$$

where N_P is the number of turns of the primary coil, N_S is the number of turns of the secondary coil, V_P is the amplitude of the alternating input voltage on the primary side of the transformer, and V_S is the amplitude of the alternating output voltage on the secondary side.

A transformer can increase, decrease, or isolate voltage, depending on the ratio of the numbers of turns in the two coils. For $N_S > N_P$, a “step-up” transformer increases the secondary voltage; for $N_S < N_P$, a “step-down” transformer decreases the secondary voltage; and for $N_S = N_P$, an “isolation” transformer produces a secondary voltage equal to the primary voltage (see later discussion). Examples of transformers are illustrated in Fig. 5-21. A key point to remember is that an alternating current is needed for a transformer to function.

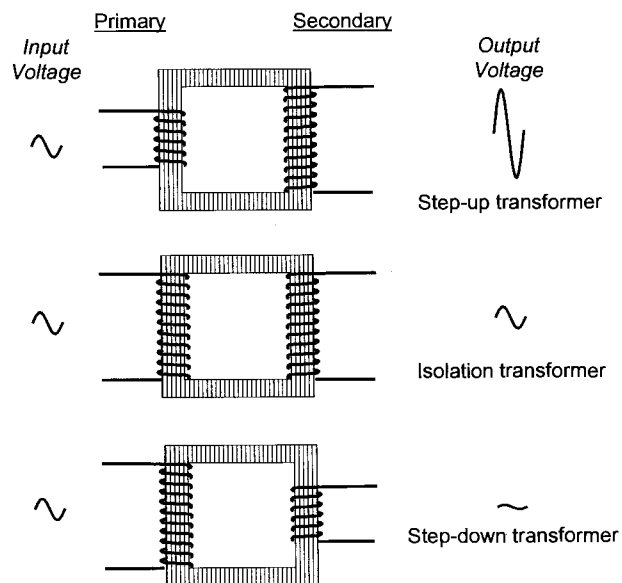


FIGURE 5-21. Transformers increase (step up), decrease (step down), or leave unchanged (isolate) the input voltage depending on the ratio of primary to secondary turns, according to the Law of Transformers. In all cases, the input and the output circuits are electrically isolated.

Power is the rate of energy production or expenditure per unit time. The SI unit of power is the watt (W), which is defined as 1 joule (J) of energy per second. For electrical devices, power is equal to the product of voltage and current:

$$P = I V \quad [5-4]$$

Because a volt is defined as 1 joule per coulomb and an ampere is 1 coulomb per second,

$$1 \text{ watt} = 1 \text{ volt} \times 1 \text{ ampere}$$

Because the power output is equal to the power input (for an ideal transformer), the product of voltage and current in the primary circuit is equal to that in the secondary circuit:

$$V_P I_P = V_S I_S \quad [5-5]$$

where I_P is the input current on the primary side and I_S is the output current on the secondary side.

Therefore, a decrease in current must accompany an increase in voltage, and vice versa. Equations 5-3 and 5-5 describe ideal transformer performance. Power losses due to inefficient coupling cause both the voltage and current on the secondary side of the transformer to be less than predicted by these equations.

Example: The ratio of primary to secondary turns is 1:1,000 in a transformer. If an input AC waveform has a peak voltage of 50 V, what is the peak voltage induced in the secondary side?

$$\frac{V_P}{V_S} = \frac{N_P}{N_S}; \frac{50}{V_S} = \frac{1}{1,000}; V_S = 50 \times 1,000 = 50,000 \text{ V} = 50 \text{ kV}$$

What is the secondary current for a primary current of 10 A?

$$V_P I_P = V_S I_S; 50 \text{ V} \times 10 \text{ A} = 50,000 \text{ V} \times I_S; I_S = 0.001 \times 10 \text{ A} = 10 \text{ mA}$$

The high-voltage section of an x-ray generator contains a step-up transformer, typically with a primary-to-secondary turns ratio of 1:500 to 1:1,000. Within this range, a tube voltage of 100 kVp requires an input peak voltage of 200 V to 100 V, respectively. The center of the secondary winding is usually connected to ground potential ("center tapped to ground"). Ground potential is the electrical potential of the earth. Center tapping to ground does not affect the maximum potential difference applied between the anode and cathode of the x-ray tube, but it limits the maximum voltage at any point in the circuit relative to ground to one half of the peak voltage applied to the tube. Therefore, the maximum voltage at any point in the circuit for a center-tapped transformer of 150 kVp is -75 kVp or +75 kVp, relative to ground. This reduces electrical insulation requirements and improves electrical safety. In some x-ray tube designs (e.g., mammography), the anode is maintained at the same potential as the body of the insert, which is maintained at ground potential. Even though this places the cathode at peak negative voltage with respect to ground, the low kVp (less than 50 kVp) used in mammography does not present a big electrical insulation problem.

Autotransformers

A simple autotransformer consists of a single coil of wire wrapped around an iron core. It has a fixed number of turns, two lines on the input side and two lines on the output side (Fig. 5-22A). When an alternating voltage is applied to the pair of input lines, an alternating voltage is produced across the pair of output lines. The Law of Transformers (Equation 5-3) applies to the autotransformer, just as it does to the standard transformer. The output voltage from the autotransformer is equal to the input voltage multiplied by the ratio of secondary to primary turns. The primary and secondary turns are the number of coil turns between the taps of the input and output lines, respectively. The autotransformer operates on the principle of self-induction, whereas the standard transformer operates on the principle of mutual induction. Self and mutual induction are described in detail in Appendix A. The standard transformer permits much larger increases or decreases in voltage, and it electrically isolates the primary from the secondary circuit, unlike the autotransformer. A switching autotransformer has a number of taps on the input and output sides (see Fig. 5-22B), to permit small incremental increases or decreases in the output voltage.

The switched autotransformer is used to adjust the kVp produced by an x-ray generator. Standard alternating current is provided to the input side of the autotransformer, and the output voltage of the autotransformer is provided to the primary side of the high-voltage transformer. Although variable resistor circuits can be

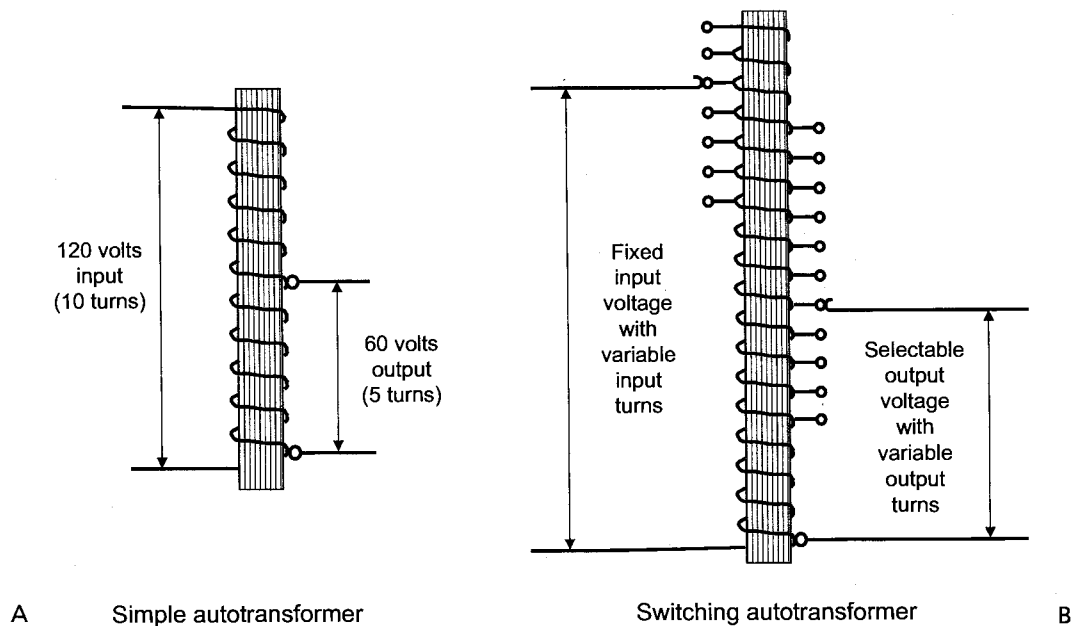


FIGURE 5-22. The autotransformer (**A**) is an iron core wrapped with a single wire. Conducting “taps” allow the ratio of input to output turns to vary, resulting in a small incremental change (increase or decrease) between input and output voltages. A switching autotransformer (**B**) allows a greater range of input to output voltages. An autotransformer does not isolate the input and output circuits from each other.

used to modulate voltage, autotransformers are more efficient in terms of power consumption and therefore are preferred.

Diodes, Triodes, Tetrodes

Diodes (having two terminals), triodes (three terminals), tetrodes (four terminals), and pentodes (five terminals) are devices that permit control of the electrical current in a circuit. The term “diode” can refer to either a solid-state or a vacuum tube device, whereas triodes, tetrodes, and pentodes usually are vacuum tube devices. These vacuum tube devices are also called “electron tubes” or “valves.” A vacuum tube diode contains two electrodes, an electron source (cathode) and a target electrode (anode), within an evacuated envelope. Diodes permit electron flow from the cathode to the anode, but not in the opposite direction. A triode contains a cathode, an anode, and a third control electrode to adjust or switch (turn on or off) the current. Tetrodes and pentodes contain additional control electrodes and function similarly to triodes.

A pertinent example of a diode is the x-ray tube itself, in which electrons flow from the heated cathode, which releases electrons by thermionic emission, to the positive anode. If the polarity of the voltage applied between the anode and cathode reverses (for example, during the alternate AC half-cycle), current flow in the circuit stops, because the anode does not release electrons to flow across the vacuum. Therefore, electrons flow only when the voltage applied between the electron source and the target electrodes has the correct polarity.

A solid-state diode contains a crystal of a *semiconductor material*—a material, such as silicon or germanium, whose electrical conductivity is less than that of a metallic conductor but greater than that of an insulator. The crystal in the diode is “doped” with trace amounts of impurity elements. This serves to increase its conductivity when a voltage is applied in one direction but to reduce its conductivity to a very low level when the voltage is applied in the opposite polarity (similar to a vacuum tube diode, described earlier). The operating principles of these solid-state devices are described in Chapter 20.

The symbols for the vacuum tube and solid-state diodes are shown in Fig. 5-23. In each case, electrons flow through the diode from the electron source (cathode) to

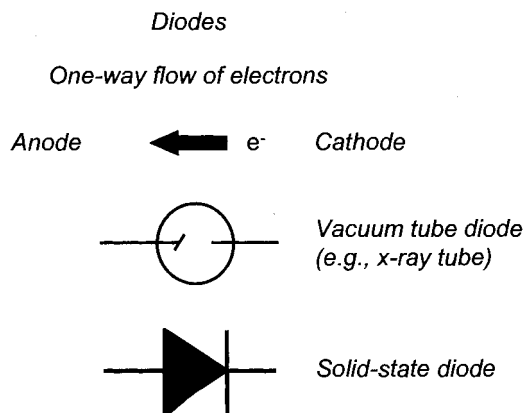


FIGURE 5-23. Diodes are electrical devices with two terminals that allow current flow in only one direction. Note that the direction of electron flow is opposite that of current flow in an electrical circuit.

the electron target (anode). The triangle in the solid-state diode symbol can be thought of as an arrow pointing in the direction *opposite* that of the electron flow. Although it is perhaps counterintuitive, electrical current is defined as the flow of positive charge (opposite the direction of electron flow). Therefore, the diode symbol points in the direction of allowed current flow through the diode.

The triode is a vacuum tube diode with the addition of a third electrode placed in close proximity to the cathode (Fig. 5-24). This additional electrode, called a *grid*, is situated between the cathode and the anode in such a way that all electrons en route from the cathode to the anode must pass through the grid structure. In the conduction cycle, a small negative voltage applied to the grid exerts a large force on the electrons emitted from the cathode, enabling on/off switching or current control. Even when a high power load is applied on the secondary side of the x-ray transformer, the current can be switched off simply by applying a relatively small negative voltage (approximately $-1,000$ V) to the grid. A “grid-switched” x-ray tube is a notable example of the triode: the cathode cup (acting as the grid) is electrically isolated from the filament structure and is used to turn the x-ray tube current on and off with microsecond accuracy. Gas-filled triodes with thermionic cathodes are known as *thyratrons*. Solid-state triodes, known as *thyristors*, operate with the use of a “gate” electrode that can control the flow of electrons through the device. Tetrodes and pentodes also control x-ray generator circuits; they have multiple grid electrodes to provide further control of the current and voltage characteristics.

Other Components

Other components of the x-ray generator (Fig. 5-25) include the high-voltage power circuit, the stator circuit, the filament circuit, the focal spot selector, and automatic exposure control circuits. In modern systems, microprocessor control and closed-loop feedback circuits help to ensure proper exposures.

Most modern generators used for radiography have automatic exposure control (AEC) circuits, whereby the technologist selects the kVp and mA, and the AEC system determines the correct exposure time. The AEC (also referred to as a *photo-*

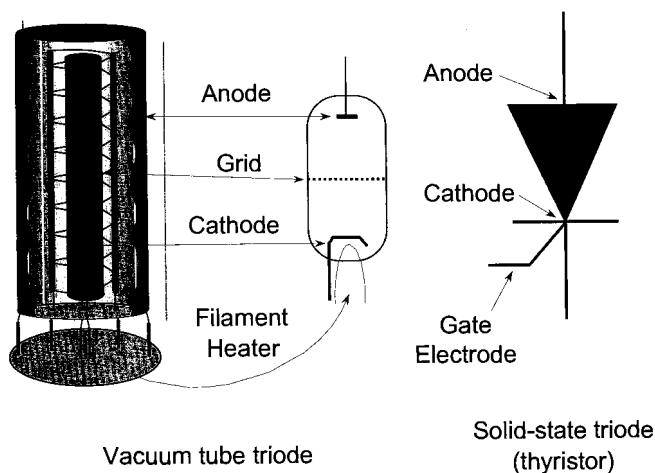


FIGURE 5-24. A diode containing a third electrode is called a triode. With a “grid” electrode, the flow of electrons between the anode and cathode can be turned on or off. The thyristor is the solid-state replacement for the tube triode.

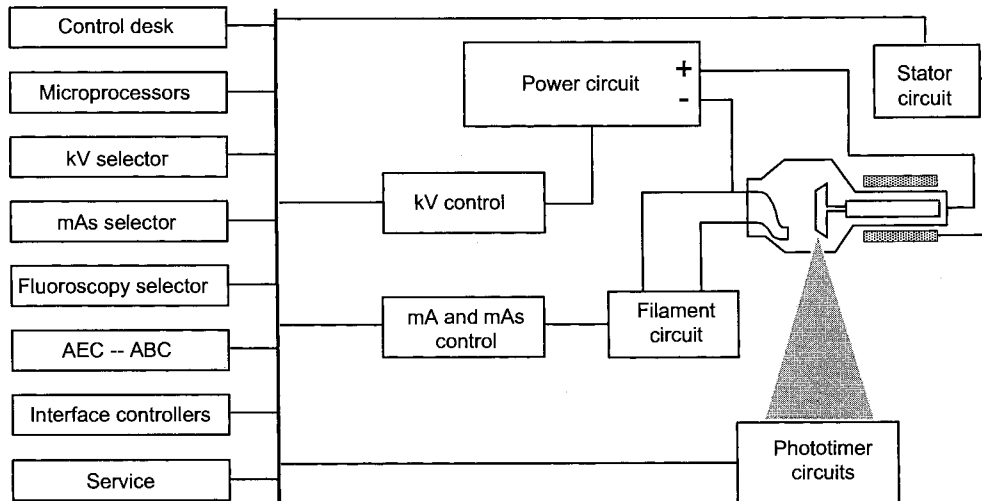


FIGURE 5-25. A modular schematic view shows the basic x-ray generator components. Most systems are now microprocessor controlled and include service support diagnostics.

timer) measures the exposure with the use of radiation detectors located near the image receptor, which provide feedback to the generator to stop the exposure when the proper exposure to the image receptor has been reached. AECs are discussed in more detail later in this chapter.

Many generators have circuitry that is designed to protect the x-ray tubes from potentially damaging overload conditions. Combinations of kVp, mA, and exposure time delivering excessive power to the anode are identified by this circuitry, and such exposures are prohibited. Heat load monitors calculate the thermal loading on the x-ray tube anode, based on kVp, mA, and exposure time, and taking into account intervals for cooling. Some x-ray systems are equipped with sensors that measure the temperature of the anode. These systems protect the x-ray tube and housing from excessive heat buildup by prohibiting exposures that would damage them. This is particularly important for CT scanners and high-powered interventional angiography systems.

Operator Console

At the operator console, the operator selects the kVp, the mA (proportional to the number of x-rays in the beam at a given kVp), the exposure time, and the focal spot size. The peak kilovoltage (kVp) determines the x-ray beam quality (penetrability), which plays a role in the subject contrast. The x-ray tube current (mA) determines the x-ray flux (photons per square centimeter) emitted by the x-ray tube at a given kVp. The product of tube current (mA) and exposure time (seconds) is expressed as milliamperere-seconds (mAs). Some generators used in radiography allow the selection of “three-knob” technique (individual selection of kVp, mA, and exposure time), whereas others only allow “two-knob” technique (individual selection of kVp and mAs). The selection of focal spot size (i.e., large or small) is usually determined by the mA setting: low mA selections allow the small focus to be used, and higher mA settings require the use of the large focus due to anode heating concerns. On

some x-ray generators, preprogrammed techniques can be selected for various examinations (i.e., chest; kidneys, ureter, and bladder; cervical spine). All console circuits have relatively low voltage and current levels that minimize electrical hazards.

5.5 X-RAY GENERATOR CIRCUIT DESIGNS

Several x-ray generator circuit designs are in common use, including single-phase, three-phase, constant potential, and medium/high-frequency inverter generators. All use step-up transformers to generate high voltage, step-down transformers to energize the filament, and rectifier circuits to ensure proper electrical polarity at the x-ray tube.

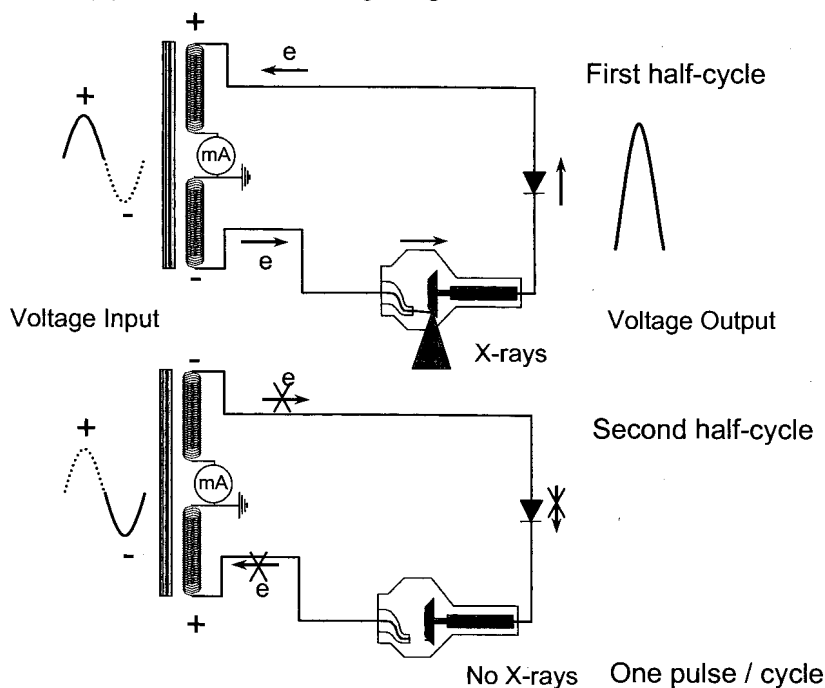
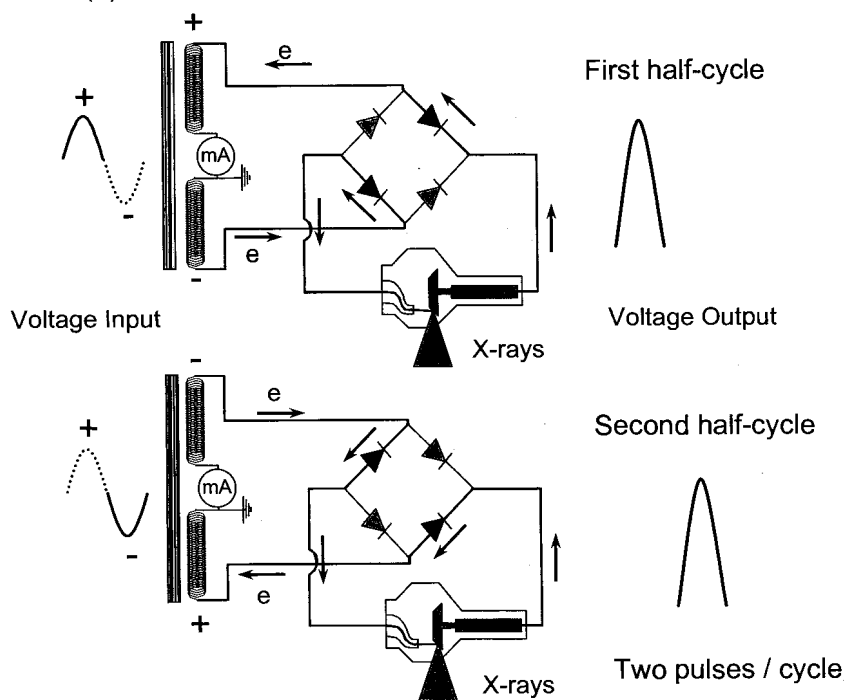
Rectifier Circuit

A rectifier is an electrical apparatus that changes alternating current into direct current. It is composed of one or more diodes. In the x-ray generator, rectifier circuits divert the flow of electrons in the high-voltage circuit so that a direct current is established from the cathode to the anode in the x-ray tube, despite the alternating current and voltage produced by the transformer. Conversion to direct current is important. If an alternating voltage were applied directly to the x-ray tube, electron back-propagation could occur during the portion of the cycle when the cathode is positive with respect to the anode. If the anode is very hot, electrons can be released by thermionic emission, and such electron bombardment could rapidly destroy the filament of the x-ray tube.

To avoid back-propagation, the placement of a diode of correct orientation in the high-voltage circuit allows electron flow during only one half of the AC cycle (when the anode polarity is positive and cathode polarity is negative) and halts the current when the polarity is reversed. As a result, a “single-pulse” waveform is produced from the full AC cycle (Fig. 5-26A), and this is called a half-wave rectified system.

Full-wave rectified systems use several diodes (a minimum of four in a bridge rectifier) arranged in a specific orientation to allow the flow of electrons from the cathode to the anode of the x-ray tube throughout the AC cycle (see Fig. 5-26B). During the first half-cycle, electrons are routed by two conducting diodes through the bridge rectifier in the high-voltage circuit and from the cathode to the anode in the x-ray tube. During the second half-cycle, the voltage polarity of the circuit is reversed; electrons flow in the opposite direction and are routed by the other two diodes in the bridge rectifier, again from the cathode to the anode in the x-ray tube. The polarity across the x-ray tube is thus maintained with the cathode negative and anode positive throughout the cycle. X-rays are produced in two pulses per cycle,

FIGURE 5-26. (a): A single-diode rectifier allows electron flow through the tube during one half of the alternating current (AC) cycle but does not allow flow during the other half of the cycle; it therefore produces x-rays with one pulse per AC cycle. **(b):** The bridge rectifier consists of diodes that reroute the electron flow in the x-ray circuit as the electrical polarity changes. The electrons flow from negative to positive polarity through two of the four diodes in the bridge rectifier circuit for the first half-cycle, and through the alternate diodes during the second half-cycle. This ensures electron flow from the cathode to the anode of the x-ray tube, producing x-rays with two pulses per AC cycle.

(a) Electron flow through *single* rectifier circuit(b) Electron flow through *bridge* rectifier circuit

independent of the polarity of the high-voltage transformer; therefore, a “two-pulse” waveform is produced with the full-wave rectified circuit.

Three-phase 6-pulse and 12-pulse generator systems also use rectifiers in the secondary circuit. In a three-phase system, multiple transformers and multiple rectifier circuits deliver direct current to the x-ray tube with much less voltage fluctuation than in a single-phase system.

Filament Circuit

The filament circuit consists of a step-down transformer, a variable resistor network (to select precalibrated filament voltage/current values), and a focal spot size selector switch to energize either the large or the small filament (see Fig. 5-27). A selectable voltage, up to about 10 V, creates a current, up to 7 A, that heats the selected filament and causes the release of electrons by thermionic emission. Calibrated resistors control the filament voltage (and therefore the filament current, and ultimately the tube current).

Single-Phase X-Ray Generator

The single-phase x-ray generator uses a single-phase input line voltage source (e.g., 220 V at 50 A), and produces either a single-pulse or a two-pulse DC waveform, depending on the high-voltage rectifier circuits. Figure 5-27 shows a diagram of the

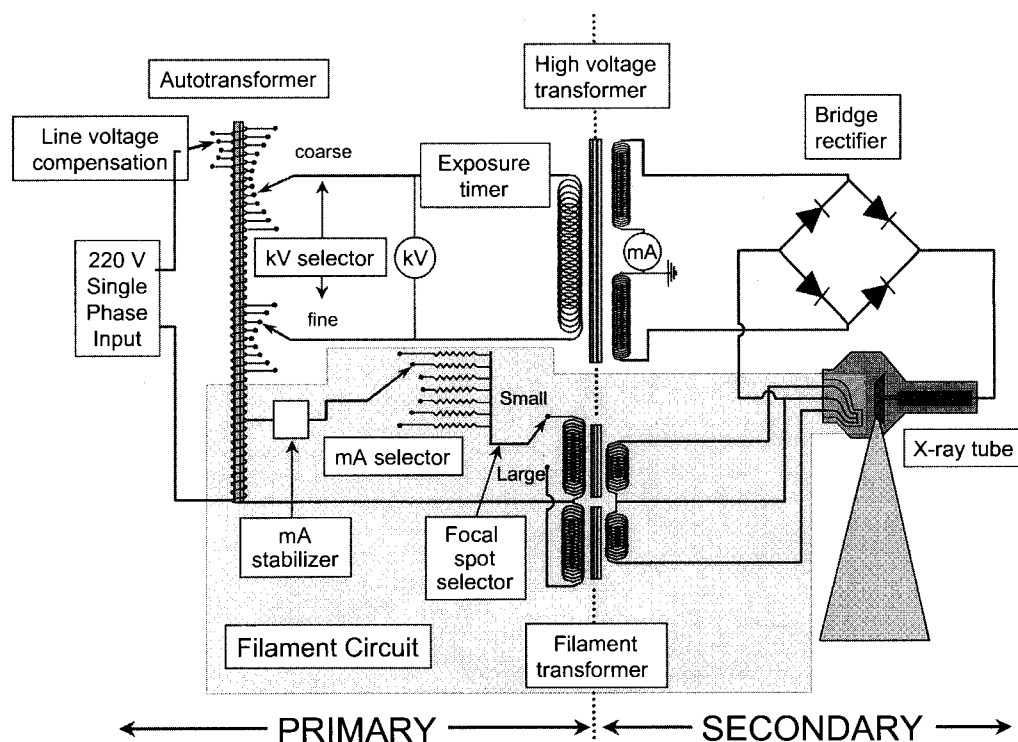


FIGURE 5-27. The diagram of a single-phase full-wave rectified (two-pulse) circuit shows the basic components and their locations in the primary or secondary side of the generator.

basic components of a single-phase two-pulse (full-wave rectified) generator circuit. The bridge rectifier routes the electron flow so that the electrons always emerge from the cathode and arrive at the anode. Figure 5-28 illustrates the voltage and current waveforms produced by this circuit. Minimum exposure time (the time of one pulse) is typically limited to intervals of $\frac{1}{120}$ second (8 msec). Simple termination of the circuit occurs as the voltage passes through zero, when contact switches can be easily opened. X-ray exposure timers calibrated in multiples of $\frac{1}{120}$ second (e.g., $\frac{1}{20}$ second, $\frac{1}{10}$ second, $\frac{1}{5}$ second) indicate a single-phase generator.

The x-ray tube current for a specific filament current is nonlinear below 40 kV due to the space charge effect (see Fig. 5-28, lower left). Cable capacitance, which is the storage and subsequent release of charge on the high-voltage cables, smooths out the variation in voltage, as shown in the lower right plot of Fig. 5-28.

Three-Phase X-Ray Generator

Three-phase x-ray generators use a three-phase AC line source (Fig. 5-29)—three voltage sources, each with a single-phase AC waveform oriented one-third of a cycle (120 degrees) apart from the other two (i.e., at 0, 120, and 240 degrees). Three separate high-voltage transformers are wired in a “delta” configuration (Fig. 5-30). A bridge rectifier on the high-voltage side in each circuit produces two pulses per cycle for each line, resulting in six pulses per cycle for a three-phase six-pulse generator. Adding another group of three transformers in a “wye” transformer configuration and additional bridge rectifiers provides a doubling of the number of pulses, or 12 pulses per cycle, in the three-phase 12-pulse generator.

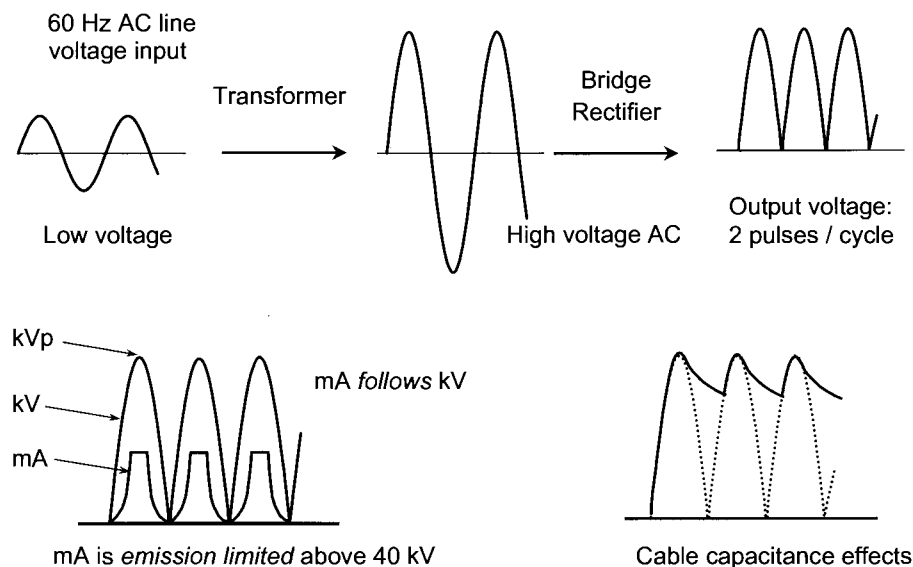


FIGURE 5-28. The single-phase full-wave rectified (two-pulse) x-ray generator uses a single-phase input voltage, produces a high-voltage alternating current (AC) waveform, and rectifies the signal, producing a pulsating direct current (DC) voltage. Variations in kV produce variations in tube current below 40 kV due to the space charge effect (**lower left**). Cable capacitance smooths the pulsating voltage variations (**lower right**).

60 Hz AC line voltage input for each of 3 phases

phase map:

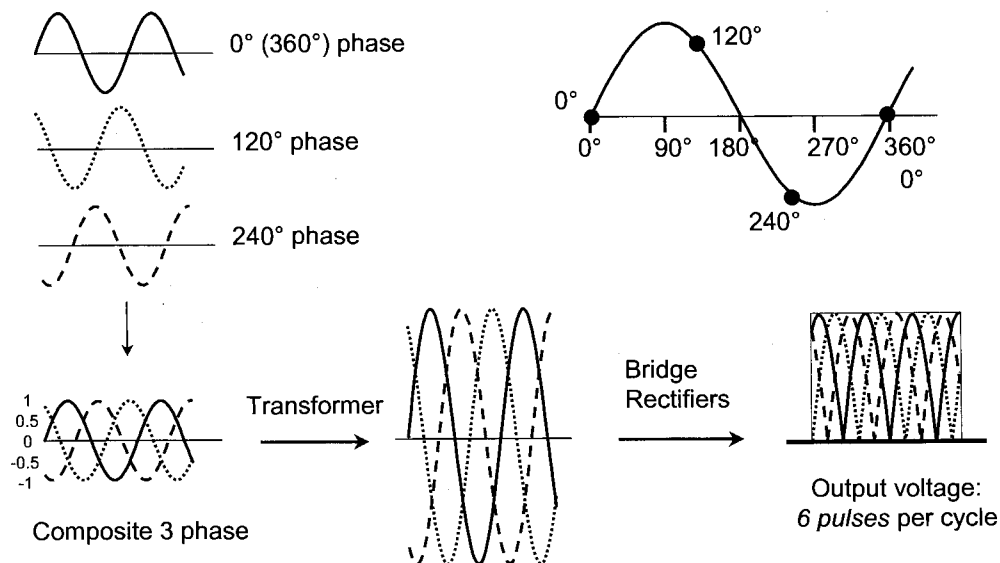


FIGURE 5-29. The three-phase line voltage source consists of three independent voltage waveforms separated by 120-degree phase increments (**top**). The three-phase voltage is transformed and rectified in the high-voltage circuit to provide a nearly constant direct current (DC) voltage with six pulses per cycle.

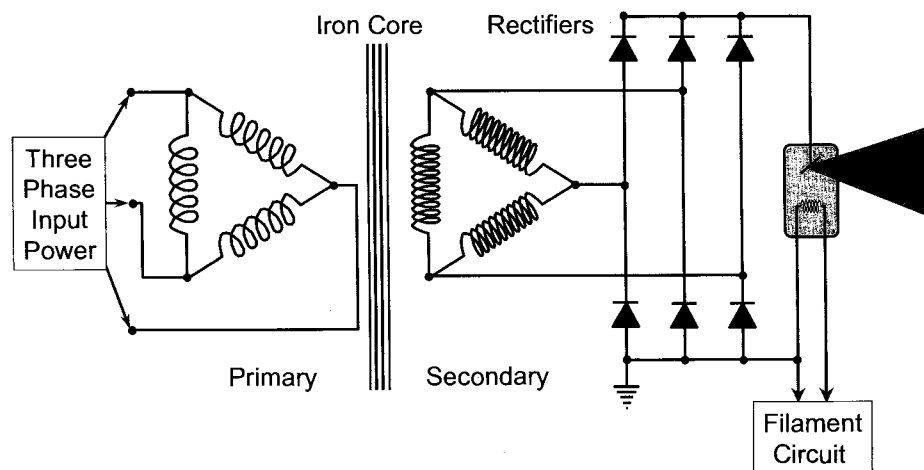


FIGURE 5-30. A simplified circuit for a three-phase six-pulse generator is illustrated. Three-phase input power is applied to the primary windings of the transformer, where each winding can be considered an independent transformer. (This is a "delta" configuration). Bridge rectification produces the high-voltage direct current (DC) output signal.

In three-phase generator designs, high-powered triode, tetrode, or pentode circuits on the secondary side of the circuit control the x-ray exposure timing. These switches turn the beam on and off during any phase of the applied voltage within extremely short times (millisecond or better accuracy). Three-phase generators deliver a more constant voltage to the x-ray tube (see Voltage Ripple) and can produce very short exposure times. However, they are more expensive than single-phase systems, are more difficult to install, have a large amount of bulky hardware and electronics, and occupy significant floor space.

Constant-Potential Generator

A constant-potential generator provides a nearly constant voltage to the x-ray tube. This is accomplished with a three-phase transformer and rectifier circuit supplying the maximum high voltage (e.g., 150 kVp) to two high-voltage electron tubes (triodes, tetrodes, or pentodes) connected in-line on the cathode side and on the anode side of the x-ray tube. These electron tubes control both the exposure duration (switching the tube voltage on and off with approximately 20- μ sec accuracy) and the magnitude of the high voltage in the secondary circuit (with 20- to 50- μ sec adjustment capability). A high-voltage comparator circuit measures the difference between the desired (set) kVp at the control console and the actual kVp on the high-voltage circuit and adjusts the grid electrodes (and thus the current and potential difference) of the electron tubes. This closed-loop feedback ensures extremely rapid kVp and mA adjustment with nearly constant voltage across the x-ray tube. The fast exposure time allows rates of 500 pulses per second or more. However, the high hardware cost, operational expense, and space requirements are major disadvantages. In all but the most demanding radiologic applications (e.g., interventional angiography), high-frequency inverter x-ray generators have replaced constant-potential generators.

High-Frequency Inverter Generator

The high-frequency inverter generator is the contemporary state-of-the-art choice for diagnostic x-ray systems. Its name describes its function, whereby a high-frequency alternating waveform (up to 50,000 Hz) is used for efficient transformation of low to high voltage. Subsequent rectification and voltage smoothing produce a nearly constant output voltage. These conversion steps are illustrated in Fig. 5-31. The operational frequency of the generator is variable, depending chiefly on the set kVp, but also on the mA and the frequency-to-voltage characteristics of the transformer.

There are several advantages of the high-frequency inverter generator. Single-phase or three-phase input voltage can be used. Closed-loop feedback and regulation circuits ensure reproducible and accurate kVp and mA values. The variation in the voltage applied to the x-ray tube is similar to that of a three-phase generator. Transformers operating at high frequencies are more efficient, more compact, and less costly to manufacture. Modular and compact design makes equipment siting and repairs relatively easy.

Figure 5-32 shows the components of a general-purpose high-frequency inverter generator. The DC power supply produces a constant voltage from either a single-phase or a three-phase input line source (the three-phase source provides

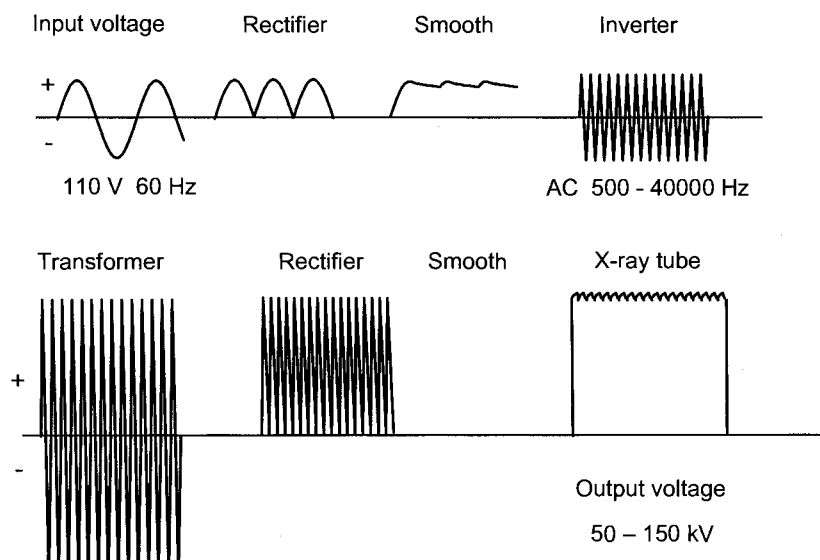


FIGURE 5-31. In a high-frequency inverter generator, a single- or three-phase alternating current (AC) input voltage is rectified and smoothed to create a direct current (DC) waveform. An inverter circuit produces a high-frequency square wave as input to the high-voltage transformer. Rectification and capacitance smoothing provide the resultant high-voltage output waveform, with properties similar to those of a three-phase system.

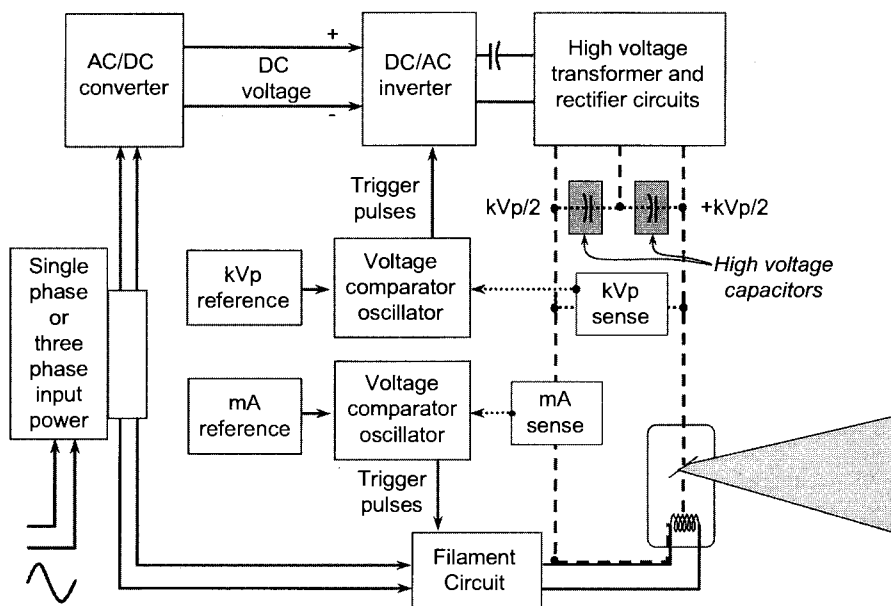


FIGURE 5-32. Modular components of the high-frequency generator, showing the feedback signals that provide excellent stability of the peak tube voltage (kVp) and tube current (mA). Solid lines in the circuit depict low voltage, dashed lines depict high voltage, and dotted lines depict low-voltage feedback signals from the sensing circuits (kVp and mA). The potential difference across the x-ray tube is determined by the charge delivered by the transformer circuit and stored on the high-voltage capacitors.

greater overall input power). Next, an inverter circuit creates the high-frequency AC waveform. This AC current supplies the high-voltage transformer and creates a waveform of fixed high voltage and corresponding low current. After rectification and smoothing, two high-voltage capacitors on the secondary circuit (shaded in Fig. 5-32) accumulate the electron charges. These capacitors produce a voltage across the x-ray tube that depends on the accumulated charge, described by the relationship $V = Q/C$, where V is the voltage (volts), Q is the charge (coulombs), and C is the capacitance (farads). During the x-ray exposure, feedback circuits monitor the tube voltage and tube current and correct for fluctuations.

For kVp adjustment, a voltage comparator/oscillator measures the difference between the reference voltage (a calibrated value proportional to the requested kVp) and the actual kVp measured across the tube by a voltage divider (the kV sense circuit). Trigger pulses generated by the comparator circuit produce a frequency that is proportional to the voltage difference between the reference signal and the measured signal. A large discrepancy in the compared signals results in a high trigger-pulse frequency, whereas no difference produces few or no trigger pulses. For each trigger pulse, the DC/AC inverter circuit produces a corresponding output pulse, which is subsequently converted to a high-voltage output pulse by the transformer. The high-voltage capacitors store the charge and increase the potential difference across the x-ray tube. When the x-ray tube potential reaches the desired value, the output pulse rate of the comparator circuit settles down to an approximately constant value, only recharging the high-voltage capacitors when the actual tube voltage drops below a predetermined limit. The feedback pulse rate (generator frequency) strongly depends on the tube current (mA), since the high-voltage capacitors discharge more rapidly with higher mA, thus actuating the kVp comparator circuit. Because of the closed-loop voltage regulation, autotransformers for kVp selection and input line voltage compensation are not necessary, unlike other generator designs.

The mA is regulated in an analogous manner to the kVp, with a resistor circuit sensing the actual mA (the voltage across a resistor is proportional to the current) and comparing it with a reference voltage. If the mA is too low, the voltage comparator/oscillator increases the trigger frequency, which boosts the power to the filament to raise its temperature and increase the thermionic emission of electrons. The closed-loop feedback circuit eliminates the need for space charge compensation circuits and automatically corrects for filament aging effects.

The high-frequency inverter generator is the preferred system for all but a few applications (e.g., those requiring extremely high power, extremely fast kVp switching, or submillisecond exposure times). In only rare instances is the constant-potential generator a better choice.

Voltage Ripple

The voltage ripple of a DC waveform is defined as the difference between the peak voltage and the minimum voltage, divided by the peak voltage and multiplied by 100:

$$\% \text{ voltage ripple} = \frac{V_{\max} - V_{\min}}{V_{\max}} \times 100 \quad [5-6]$$

A comparison of the voltage ripple for various x-ray generators is shown in Fig. 5-33. In theory, a single-phase generator, whether one-pulse or two-pulse output,

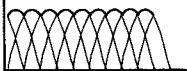
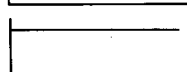
Generator type	Typical voltage waveform	kV ripple
Single-phase 1-pulse (self rectified)		100%
Single-phase 2-pulse (full wave rectified)		100%
3-phase 6-pulse		13% - 25%
3-phase 12-pulse		3% - 10%
Medium-high frequency inverter		4% - 15%
Constant Potential		<2%

FIGURE 5-33. Typical voltage ripple for various x-ray generators used in diagnostic radiology varies from 100% voltage ripple for a single-phase generator to almost 0% ripple for a constant-potential generator.

has 100% voltage ripple. The actual voltage ripple for a single-phase generator is less than 100% because of cable capacitance effects (longer cables produce a greater capacitance). In effect, capacitance “borrows” voltage from the cable and returns it a short time later, causing the peaks and valleys of the voltage waveform to be smoothed out. Three-phase 6-pulse and 12-pulse generators have a voltage ripple of 3% to 25%. High-frequency generators have a ripple that is kVp and mA dependent, typically similar to a three-phase generator, ranging from 4% to 15%. Higher kVp and mA settings result in less voltage ripple for these generators. Constant-potential generators have an extremely low ripple (approximately 2%), which is essentially considered DC voltage.

5.6 TIMING THE X-RAY EXPOSURE IN RADIOGRAPHY

Electronic Timers

Digital timers have largely replaced electronic timers based on resistor-capacitor circuits and charge-discharge times. These digital timer circuits have extremely high reproducibility and microsecond accuracy. The timer activates and terminates a switch, on either the primary or the secondary side of the x-ray tube circuit, with a low-voltage signal. Precision and accuracy of the x-ray exposure time depends chiefly on the type of exposure switch employed in the x-ray system. X-ray systems also use a countdown timer (known as a backup timer) to terminate the exposure in the event of timer or exposure switch failure.

Switches

Mechanical contactor switches are used solely in single-phase, low-power x-ray generators. An electrical circuit controlled by the timer energizes an electromagnet con-

nected to the contactors that close the circuit on the low-voltage side of the transformer. After the selected time has elapsed, the electromagnet is de-energized to allow the contactors to open and turn off the applied voltage to the x-ray tube. When the circuit has significant power (high positive or negative voltage), the switch contacts will remain in the closed position. Only when the power in the circuit passes through zero, at intervals of $1/120$ second (8 msec), can the contactors release to terminate the exposure. The accuracy and reproducibility of these timers are poor, especially for short exposure times, because the increment of time occurs in 8-msec intervals.

For three-phase and constant-potential generators, high-voltage triode, tetrode, or pentode switches initiate and terminate the exposure on the secondary side of the circuit under the control of the electronic timer or phototimer. The operating principles of these devices were discussed earlier (see Diodes, Triodes, Tetrodes). Because the ability to open and close the switch is independent of the power on the circuit, exposure times of 1 msec or less are possible.

The high-frequency inverter generator typically uses electronic switching on the primary side of the high-voltage transformer to initiate and stop the exposure. A relatively rapid response resulting from the high-frequency waveform characteristics of the generator circuit allows exposure times as short as 2 msec.

Alternatively, a grid-controlled x-ray tube can be used with any type of generator to switch the exposure on and off by applying a bias voltage (about ~2000 V) to the focusing cup.

Phototimers

Phototimers are often used instead of manual exposure time settings in radiography. Phototimers operate by measuring the actual amount of radiation incident on the image receptor (e.g., screen-film detector) and terminating x-ray production when the proper amount is obtained. Thus the phototimer helps provide a consistent exposure to the image receptor by compensating for thickness and other variations in attenuation in a particular patient and from patient to patient.

The phototimer system consists of one or more radiation detectors, an amplifier, a density selector, a comparator circuit, a termination switch, and a backup timer. X-rays transmitted through the patient instantaneously generate a small signal in an ionization chamber (or other radiation-sensitive device such as a solid-state detector). An amplifier boosts the signal, which is fed to a voltage comparator and integration circuit. When the accumulated signal equals a preselected reference value, an output pulse terminates the exposure. A user-selectable “density” switch on the generator control panel adjusts the reference voltage to modify the exposure.

For most general diagnostic radiography, the phototimer sensors are placed in front of the image receptor to measure the transmitted x-ray flux through the patient (see Figure 5-34). Positioning in front of the image receptor is possible because of the high transparency of the ionization sensor at higher kVp values (although this is not the case for mammography). In the event of a phototimer detector or circuit failure, a “backup timer” safety device terminates the x-ray exposure after a preset time.

In order to compensate for different x-ray penetration through distinctly different anatomy, three sensors are typically used in a chest or table cassette stand with phototiming capability as shown in Fig. 5-34. The technologist can select

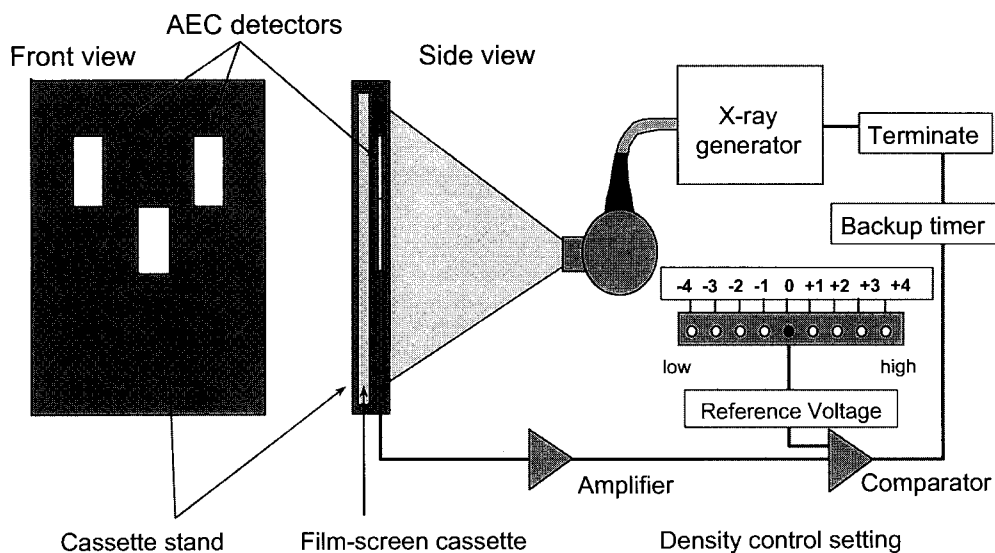


FIGURE 5-34. Automatic exposure control (AEC) detectors measure the radiation incident on the detector and terminate the exposure according to a preset film exposure time. A front view (**left**) and side view (**middle**) of a chest cassette stand and the location of ionization chamber detectors are shown. Density control selection is available at the operator console of the generator.

which photocells are used for each radiographic application. For instance, in posteroanterior chest imaging, the two outside chambers are usually activated, and the transmitted x-ray beam through the lungs determines the exposure time. This prevents lung “burnout,” which occurs when the transmitted x-ray flux is measured under the highly attenuating mediastinum. AEC is a synonym for phototiming.

Falling-Load Generator Control Circuit

The falling-load generator circuit works in concert with the phototiming (AEC) subsystem. “Falling-load” capability delivers the *maximum* possible mA for the selected kVp by considering the instantaneous heat load characteristics of the x-ray tube. Preprogrammed electronic circuitry determines the maximal power load limit of the x-ray anode and then continuously reduces the power as the exposure continues (Fig. 5-35). This delivers the desired amount of radiation to the image receptor in the shortest possible exposure time. For example, assume that a technique of 50 mAs at 80 kVp delivers the proper exposure to the image receptor. The single-exposure tube rating chart (described later) indicates that 400 mA is the maximum *constant* mA that is tolerated by the focal spot. To deliver 50 mAs requires an exposure time of 125 msec. A falling-load generator, however, initially provides a tube current of 600 mA and decreases the tube current continuously according to a predetermined power load curve. The initial higher mA with falling load delivers the 50 mAs exposure in 100 msec, 25 msec less than the exposure time for constant tube current. The reduced exposure time decreases possible patient motion artifacts.

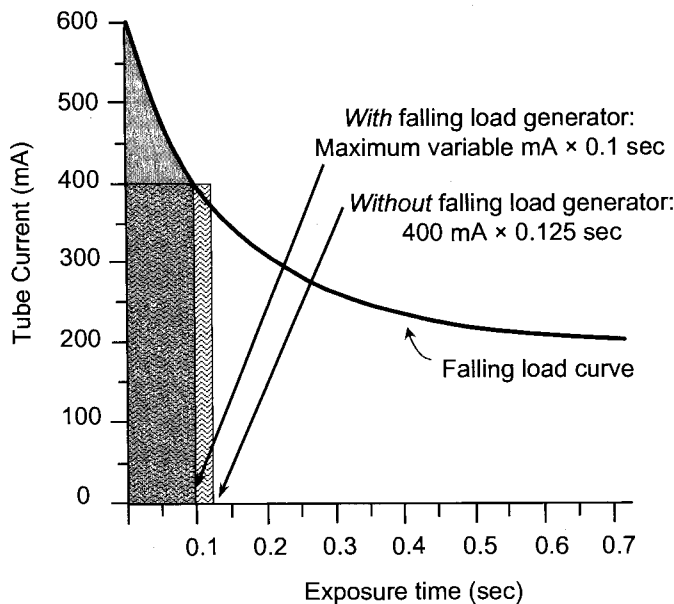


FIGURE 5-35. The falling-load generator control circuit provides a continuously decreasing mA with time, in order to deliver the maximal continuous x-ray output possible within power deposition capabilities of the focal spot. This generator circuit is commonly used in conjunction with automatic exposure control (AEC) devices to provide the shortest exposure time possible for a given kVp and mAs technique. The example in the figure shows that, for a 50-mAs exposure, the falling-load generator can deliver the same radiation in 0.10 second that would require 0.125 second with a constant tube current.

5.7 FACTORS AFFECTING X-RAY EMISSION

The output of an x-ray tube is often described by the terms quality, quantity, and exposure. *Quality* describes the penetrability of an x-ray beam, with higher energy x-ray photons having a larger HVL and higher “quality.” *Quantity* refers to the number of photons comprising the beam. *Exposure*, defined in Chapter 3, is nearly proportional to the energy fluence of the x-ray beam and therefore has quality and quantity associated characteristics. X-ray production efficiency, exposure, quality, and quantity are determined by six major factors: x-ray tube target material, voltage, current, exposure time, beam filtration, and generator waveform.

1. The target (anode) material affects the *efficiency* of bremsstrahlung radiation production, with output exposure roughly proportional to atomic number. Incident electrons are more likely to have radiative interactions in higher- Z materials (see Equation 5-1). The energies of characteristic x-rays produced in the target depend on the target material. Therefore, the target material affects the quantity of bremsstrahlung photons and the quality of the characteristic radiation.
2. Tube voltage (kVp) determines the maximum energy in the bremsstrahlung spectrum and affects the quality of the output spectrum. In addition, the efficiency of x-ray production is directly related to tube voltage. Exposure is approximately proportional to the square of the kVp in the diagnostic energy range:

$$\text{Exposure} \propto \text{kVp}^2 \quad [5-7]$$

For example, according to Equation 5-7, the relative exposure of a beam generated with 80 kVp compared with that of 60 kVp for the same tube current and exposure time is calculated as follows:

$$\left(\frac{80}{60}\right)^2 \cong 1.78$$

Therefore, the output exposure increases by approximately 78% (Fig. 5-36). *An increase in kVp increases the efficiency of x-ray production and the quantity and quality of the x-ray beam.*

Changes in the kVp must be compensated by corresponding changes in mAs to maintain the same exposure. At 80 kVp, 1.78 units of exposure occur for every 1 unit of exposure at 60 kVp. To achieve the original 1 unit of exposure, the mAs must be adjusted to $1/1.78 = 0.56$ times the original mAs, which is a *reduction* of 44%. An additional consideration of technique adjustment concerns the x-ray attenuation characteristics of the patient. To achieve equal transmitted exposure through a typical patient (e.g., 20 cm tissue), the mAs varies with the fifth power of the kVp ratio:

$$\left(\frac{kVp_1}{kVp_2}\right)^5 \times mAs_1 = mAs_2 \quad [5-8]$$

According to Equation 5-8, if a 60-kVp exposure requires 40 mAs for a proper exposure through a typical adult patient, at 80 kVp the proper mAs is approximately

$$\left(\frac{60}{80}\right)^5 \times 40 \text{ mAs} \cong 9.5 \text{ mAs}$$

or about one fourth of the original mAs. The value of the exponent (between four and five) depends on the thickness and attenuation characteristics of the patient.

3. The tube current (mA) is equal to the number of electrons flowing from the cathode to the anode per unit time. The exposure of the beam for a given kVp and filtration is proportional to the tube current.
4. The exposure time is the duration of x-ray production. The quantity of x-rays is directly proportional to the product of tube current and exposure time (mAs).
5. The beam filtration modifies the quantity and quality of the x-ray beam by selectively removing the low-energy photons in the spectrum. This reduces the pho-

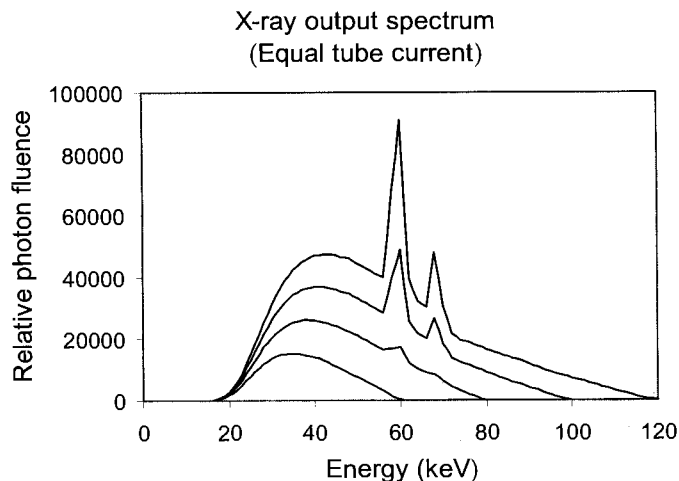


FIGURE 5-36. X-ray tube output intensity varies strongly with tube voltage (kVp). In this example, the same tube current and exposure times (mAs) are compared for 60 to 120 kVp. The relative area under each spectrum roughly follows a squared dependence (characteristic radiation is ignored).

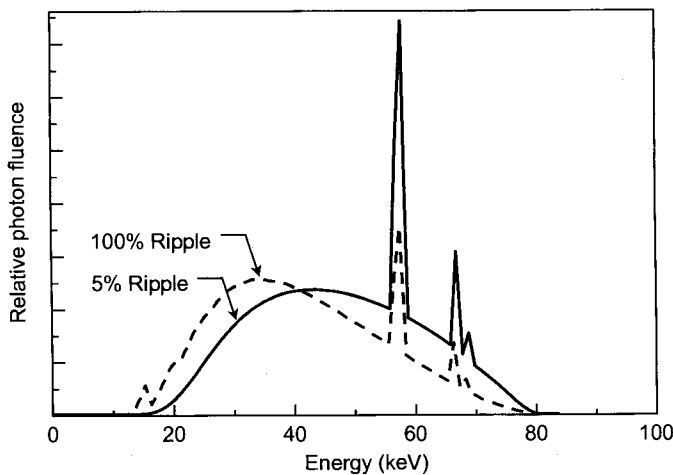


FIGURE 5-37. Output intensity (bremsstrahlung) spectra for the same tube voltage (kVp) and the same tube current and exposure time (mAs) demonstrate the higher effective energy and greater output of a three-phase or high-frequency generator voltage waveform (approximately 5% voltage ripple), compared with a single-phase generator waveform (100% voltage ripple).

ton number (quantity) and shifts the average energy to higher values, increasing the quality.

6. The generator waveform affects the quality of the emitted x-ray spectrum. For the same kVp, a single-phase generator provides a lower average potential difference than does a three-phase or high-frequency generator. Both the quality and quantity of the x-ray spectrum are affected (Fig. 5-37).

In summary, the x-ray *quantity* is approximately proportional to $Z_{\text{target}} \times \text{kVp}^2 \times \text{mAs}$. The x-ray *quality* depends on the kVp, the generator waveform, and the tube filtration. Exposure depends on both the quantity and quality of the x-ray beam. Compensation for changes in kVp with radiographic techniques requires adjustments of mAs on the order of the fifth power of the kVp ratio, because kVp determines quantity, quality, and transmission through the object, whereas mAs determines quantity only.

5.8 POWER RATINGS AND HEAT LOADING

Power Ratings of Generators and X-Ray Tube Focal Spots

The *power rating* describes the energy per unit time that can be supplied by the x-ray generator or received by the x-ray tube during operation. As mentioned earlier, the power delivered by an electric circuit is equal to the product of the voltage and the current, where $1 \text{ W} = 1 \text{ V} \times 1 \text{ A}$. For an x-ray generator or an x-ray tube focal spot, the power rating in kilowatts (kW) is the average power delivered by the maximum tube current (A_{max}) for 100 kVp and 0.1-second exposure time:

$$\text{Power} = 100 \text{ kVp} \times A_{\text{max}} \text{ for a 0.1-second exposure} \quad [5-9]$$

For an x-ray tube focal spot, this power rating can be determined using a tube rating chart. The maximal tube current for high-powered generators can exceed 1,000 mA (1 A) for short exposures. Large focal spots, high anode rotation speeds, massive anodes (providing large heat sinks), small anode angles, and large anode diameters contribute to high power ratings. The *average* kV, which depends on the

TABLE 5-4. ROOT-MEAN-SQUARE VOLTAGE (V_{rms}) AS A FRACTION OF PEAK VOLTAGE (V_{peak})

Generator Type	V_{rms} as a fraction of V_{peak}
Single-phase	0.71
Three-phase six-pulse	0.95
Three-phase 12-pulse	0.99
High-frequency	0.95–0.99
Constant-potential	1.00

generator waveform and cable capacitance properties, is used in the equation to determine the power delivered over the 100-msec period. For all but constant-potential generators, the average kV is less than the kVp (due to voltage ripple), and it is substantially less for a single-phase generator. Exact determination of the power rating requires multiplication by the “root-mean-square” voltage (V_{rms}), as listed in Table 5-4 for various x-ray generator voltage waveforms. (V_{rms} is the constant voltage that would deliver the same power as the varying voltage waveform). This correction is usually ignored for all but the single-phase generator.

Example 1: A high-frequency generator is capable of producing a maximum of 600 mA at 100 kVp for a 100-msec exposure. According to Equation 5-9, the power rating of the generator is $100 \text{ kVp} \times 0.6 \text{ A} = 60 \text{ kW}$.

Example 2: An x-ray focal spot is limited to 200 mA when the generator is operated at 100 kVp for 0.1 second. The power rating of the focal spot is $100 \text{ kVp} \times 0.2 \text{ A} = 20 \text{ kW}$.

These two examples illustrate the power delivery capability of the generator and power deposition tolerance of the focal spot. Table 5-5 classifies the x-ray generator in terms of power output, and Table 5-6 lists typical x-ray tube focal spot power ratings. X-ray generators have circuits to prohibit combinations of kVp, mA, and exposure time that will exceed the single-exposure power deposition tolerance of the focal spot.

TABLE 5-5. X-RAY GENERATOR USAGE AND POWER RATINGS

Generator Type	Power Level Usage	Applications	Typical Power Requirements (kW)
Single-pulse, one-phase	Very low	Dental and handheld fluoroscopy units	≤ 2
Two-pulse, one-phase	Low to medium	General fluoroscopy and radiography	≥ 10 and ≤ 50
Six-pulse, three-phase	Medium to high	Light special procedures	≥ 30 and ≤ 100
Twelve-pulse, three-phase	High	Interventional and cardiac angiography	≥ 50 and ≤ 150
Constant-potential	High	Interventional and cardiac angiography	≥ 80 and ≤ 200
High-frequency	All ranges	All radiographic procedures	≥ 2 and ≤ 150

TABLE 5-6. X-RAY TUBE FOCAL SPOT SIZE AND TYPICAL POWER RATING

Nominal X-ray Tube Focal Spot Size (mm)	Typical Power Rating (kW)
1.2–1.5	80–125
0.8–1.0	50–80
0.5–0.8	40–60
0.3–0.5	10–30
0.1–0.3	1–10
<0.1 (micro-focus tube)	<1

Heat Loading

The Heat Unit

The heat unit (HU) provides a simple way of expressing the energy deposition on and dissipation from the anode of an x-ray tube. Calculating the number of HUs requires knowledge of the parameters defining the radiographic technique:

$$\text{Energy (HU)} = \text{peak voltage (kVp)} \times \text{tube current (mA)} \times \text{exposure time (sec)} \quad [5-10]$$

Although this expression is correct for single-phase generator waveforms, the HU underestimates the energy deposition of three-phase, high-frequency, and constant-potential generators because of their lower voltage ripple and higher average voltage. A multiplicative factor of 1.35 to 1.40 compensates for this difference, the latter value being applied to constant-potential waveforms. For example, an exposure of 80 kVp, 250 mA, and 100 msec results in the following HU deposition on the anode:

Single-phase generators: $80 \times 250 \times 0.1 = 2,000$ HU

Three-phase generators: $80 \times 250 \times 0.1 \times 1.35 = 2,700$ HU

Constant-potential generators: $80 \times 250 \times 0.1 \times 1.40 = 2,800$ HU

For continuous x-ray production (fluoroscopy), the HU/sec is defined as follows:

$$\text{HU/sec} = \text{kVp} \times \text{mA} \quad [5-11]$$

For other than single-phase operation, the heat input rate is adjusted by the multiplicative factors mentioned previously. Heat accumulates by energy deposition and simultaneously disperses by anode cooling. The anode cools faster at higher temperatures, so that for most fluoroscopy procedures a steady-state equilibrium exists. This is taken into consideration by the anode thermal characteristics chart for heat input and heat output (cooling) curves; otherwise heat deposition for a given fluoroscopy time would be overestimated.

The Joule

The joule (J) is the SI unit of energy. One joule is deposited by a power of one watt acting for one second ($1 \text{ J} = 1 \text{ W} \times 1 \text{ sec}$). The energy, in joules, deposited in the anode is calculated as follows:

$$\text{Energy (J)} = \text{Root-mean-square voltage (V)} \times \text{Tube current (A)} \times \text{Exposure time (sec)} \quad [5-12]$$

As mentioned earlier, the root-mean-square voltage is the constant voltage that would deliver the same power as the varying voltage waveform. Table 5-4 lists the x-ray generator peak voltage and V_{rms} . The relationship between the joule and the heat unit can be approximated as follows:

$$\text{Heat input (HU)} \cong 1.4 \times \text{Heat input (J)} \quad [5-13]$$

5.9 X-RAY EXPOSURE RATING CHARTS

Rating charts exist to determine operational limits of the x-ray tube for single and multiple exposures and the permissible heat load of the anode and the tube housing. The single-exposure chart contains the information to determine whether a proposed exposure is possible without causing tube damage. On systems capable of multiple exposures (e.g., angiographic units), a multiple-exposure rating chart is also necessary. For continuous operation (e.g., fluoroscopy), the anode heat input and cooling chart and the housing heat input and cooling chart are consulted. These charts show the limitations and allowable imaging techniques for safe operation. A number of system parameters influence the characteristics and values of the rating charts, including (a) x-ray tube design (focal spot size, anode rotation speed, anode angle, anode diameter, anode mass, and use of heat exchangers) and (b) x-ray generator type (single-phase, three-phase, high-frequency, or constant-potential). A rating chart is specific to a particular x-ray tube and must not be used for other tubes.

Single-Exposure Rating Chart

A single-exposure rating chart provides information on the allowed combinations of kVp, mA, and exposure time (power deposition) for a particular x-ray tube, focal spot size, anode rotation speed, and generator type (no accumulated heat on the anode). For each tube, there are usually four individual charts with peak kilovoltage as the y-axis and exposure time as the x-axis, containing a series of curves, each for a particular mA value. Each of the four charts is for a specific focal spot size and anode rotation speed. Each curve represents the maximal allowable tube current for a particular kVp and exposure time. Figure 5-38 shows a typical set of single-exposure rating charts, with each graph representing a specific focal spot size (or focal spot power rating in kW) and anode rotation speed (or the stator frequency in Hz). Occasionally, there are rating charts of mA versus time, which contain a series of curves, each for a particular kVp.

The mA curves in any of the individual charts in Fig. 5-38 have larger values on the lower left of the graph and smaller values toward the upper right. This indicates the ability to use higher mA values for short-exposure, low-kVp techniques and implies that lower mA values may be safely used. It also shows that only low mA values are possible with long exposure times and/or high-kVp techniques. Along a given mA curve, as the exposure time increases, the kVp decreases, to limit the power deposited during the exposure. Comparison of the rating charts for the large focal spot and the small focal spot for the same anode rotation speed in Fig. 5-38 shows that the large focal spot allows a larger power loading.

There are many ways to determine whether a given single-exposure technique is allowed. The mA curves on the graph define the transition from allowed expo-

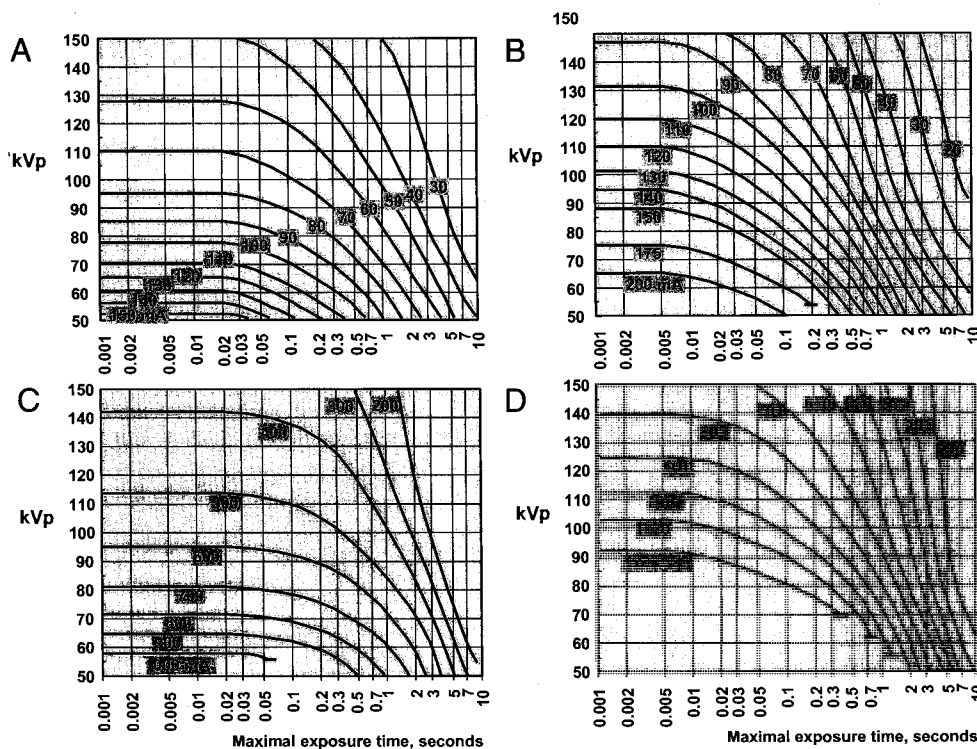


FIGURE 5-38. The single exposure rating charts are for the following focal spot and anode rotation speeds: **(A)** 0.3 mm focal spot, 10 kW power rating, 3,000 rpm anode rotation speed; **(B)** 0.3 mm, 10 kW, 10,000 rpm; **(C)** 1.2 mm, 120 kW, 3,000 rpm; **(D)** 1.2 mm, 120 kW, 10,000 rpm. Each curve plots peak kilovoltage (kVp) on the vertical axis versus time (seconds) on the horizontal axis, for a family of tube current (mA) curves indicating the maximal power loading.

tures to disallowed exposures for a specified kVp and exposure time. One method is as follows:

1. Find the intersection of the requested kVp and exposure time.
2. Determine the corresponding mA. If necessary, estimate by interpolation of adjacent mA curves. This is the *maximal* mA allowed by the tube focal spot.
3. Compare the *desired* mA to the *maximal* mA allowed. If the desired mA is larger than the maximal mA, the exposure is *not* allowed. If the desired mA is equal to or smaller than the maximal mA, the exposure *is* allowed.

For mA versus time plots with various kVp curves, the rules are the same but with a simple exchange of kVp and mA labels.

Example: Determine whether the following exposures are allowed according to the charts in Fig. 5-38:

- (a) 75 kVp, 100 mA, 0.125-sec exposure, 10 kW focal spot, 10,000 rpm anode rotation speed.
- (b) Same as in (a) except with 3,000 rpm anode rotation speed.
- (c) 120 kVp, 700 mA, 0.05-sec exposure, 100 kW focal spot, 10,000 rpm.

Answers: (a) Yes, (b) no, (c) yes.

Previous exposures must also be considered when deciding whether an exposure is permissible, because there is also a limit on the total accumulated heat capacity of the anode. The anode heat input and cooling charts must be consulted in this instance.

Multiple-Exposure Rating Chart

Multiple-exposure rating charts are used in angiography and in other imaging applications for which sequential exposures are necessary. These charts can be helpful in the planning of an examination, but they are used infrequently because of the variations in imaging requirements that are necessary during an examination. Most modern angiographic equipment includes a computerized heat loading system to indicate the anode status and determine the safety of a given exposure or series of exposures. Similarly, in CT, the computer system monitors the x-ray tube exposure sequence. When the tube heat limit has been reached (for either the anode loading or the housing loading), further image acquisition is prevented until the anode (or housing) has cooled sufficiently to ensure safety. In modern microprocessor-controlled x-ray units, the x-ray tube heat limitations are programmed into the software. This has made tube rating charts less important.

Anode Heat Input and Cooling Chart

Multiple x-ray exposures and continuous fluoroscopy deliver a cumulative heat energy deposition on the anode. Even if a single-exposure technique is allowed, damage to the anode could still occur if there is a large amount of stored heat from previous exposures. The *anode cooling chart* shows the remaining heat load of the anode (usually in HU but sometimes also in joules) versus time as the anode cools (Fig. 5-39). The anode cooling curve is steeper at larger heat loads because higher temperatures cause more rapid cooling (the “black-body” effect—cooling rate $\propto T^4$, for temperature T). The maximum anode heat load is the upper value on the y-axis

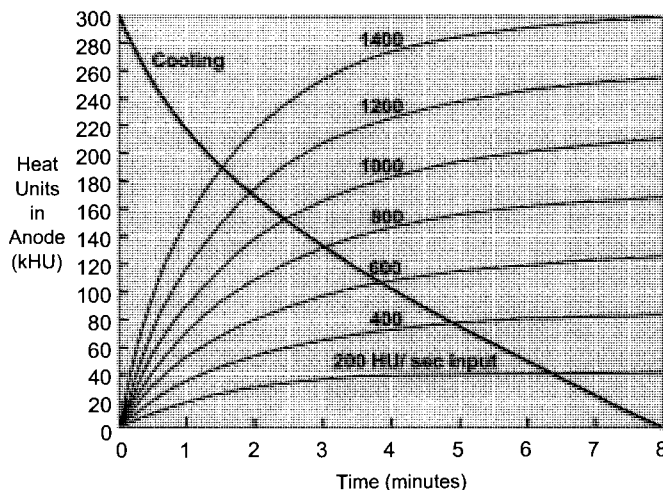


FIGURE 5-39. Anode thermal characteristics chart plots heat units (kHU) on the vertical axis and time (minutes) on the horizontal axis. Each curve labeled with a number depicts anode heat loading as a function of beam-on time for a particular heat deposition rate (HU/sec). The cooling curve shows the rate of cooling and indicates faster cooling with higher temperatures. The maximum anode heat loading is the largest value on the vertical axis.

of the chart. Like the single-exposure rating charts, the anode heat input and cooling chart is specific to a particular x-ray tube.

How is the anode cooling chart used? After a series of exposures, the total heat load accumulated on the anode is calculated as the sum of the HU incident (with adjustment factors applied) per exposure. If the radiographic technique is high or a large number of sequential exposures have been taken, it might be necessary to wait before reusing the x-ray tube in order to avoid anode damage. The cooling chart specifies how long to wait.

Example: How long must one wait to initiate a sequence of 20 films, each acquired at 85 kVp, 80 mAs, if there are 250,000 HU accumulated on the tube anode from a previous acquisition?

Answer: The maximal heat load of the anode, according to Fig. 5-39, is 300,000 HU. The 20 exposures will result in a total of (85×80) HU per film, for a total of 136,000 HU. With 250,000 HU already on the anode, the total would be 386,000 HU, which exceeds the maximum by 86,000 HU. Therefore, knowledge of the time required to dissipate 86,000 HU is necessary. This is determined from the curve by starting at 250,000 HU (0.6 min) and proceeding along the curve until 86,000 HU has been eliminated, to 164,000 HU (2.2 min). The time between the two points on the x-axis is the required waiting time ($2.2 - 0.6 = 1.6$ min). If the anode were not as hot, it would take longer to eliminate the same amount of heat.

On the same chart is a collection of heat input curves, depicted by an HU/sec or joule/sec (watt) value that corresponds to the continuous heat input resulting from fluoroscopy operation. These curves initially rise very fast but reach a plateau after a short time (high dose fluoro techniques are an exception). At this point, the rate of heat energy input into the anode equals the rate of heat energy dissipated by radiative emission. For standard levels of fluoroscopy, then, exposure time is essentially unlimited with respect to anode heat loading. The heat input curves are useful for determining the amount of accumulated heat on the anode after a given amount of fluoroscopy time, taking into account the simultaneous cooling.

Example: Determine how much heat accumulates on the anode from a room temperature start using 7 minutes of fluoroscopy (80 kVp and 5 mA).

Answer: The number of HU/sec is calculated by the product of kVp and mA: $80 \text{ kVp} \times 5 \text{ mA} = 400 \text{ HU/sec}$ input. This curve is identified on the anode heat input chart (400 HU/sec). Room temperature implies 0 HU initially. The total number of HU accumulated on the anode is given by the height of this curve at 7 minutes. It is approximately 82,000 HU. Note that if heat input did not follow the curve, 168,000 HU would have accumulated on the anode. Anode cooling occurs simultaneous with heating, and it is more rapid for a hotter anode.

X-ray tube vendors now manufacture high-powered x-ray tubes that exceed 5 MHU anode heat loading with extremely high cooling rates for CT devices, achieved through advanced anode, rotor bearing, and housing designs.

Housing Cooling Chart

The heat generated in the anode eventually transfers to the tube housing, and the temperature of the insulating oil and external housing assembly increases with use.

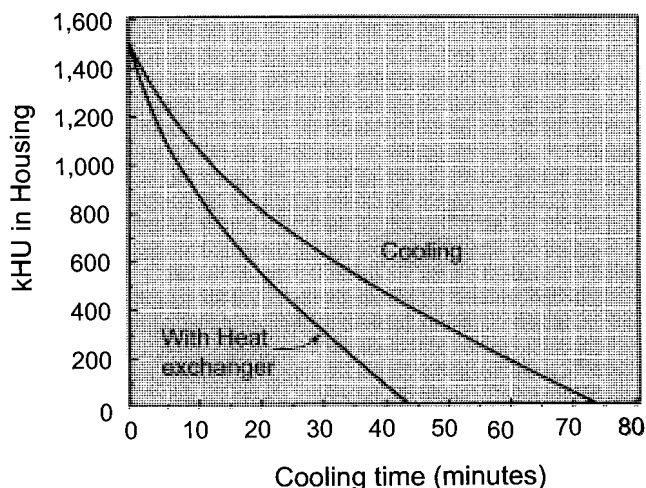


FIGURE 5-40. The tube housing cooling curves depict cooling with and without a heat exchanger. The maximal heat loading of the tube housing is the largest value on the vertical axis.

An expansion bellows in the tube housing (see Fig. 5-6) allows for the safe heating and expansion of the volume of insulation oil during tube operation. If the oil is heated to a dangerous temperature, a microswitch attached to the expanded bellows prohibits further exposures. Cooling of the housing is more efficient if a heat exchanger system is used to circulate oil through an external heat radiator or to pump cooled water through a water jacket in the housing. In either case, the enhanced removal of heat makes it less likely that the housing limit will be exceeded. The housing cooling curves in Fig. 5-40 show this behavior. The tube housing heat load typically exceeds that of the anode, and heat load ratings of 1.5 MHU and higher are common.

In summary, x-rays are the basic radiologic tool for a majority of medical diagnostic imaging procedures. Knowledge of x-ray production, x-ray generators, and x-ray beam control is important for further understanding of the image formation process and the need to obtain the highest image quality at the lowest possible radiation exposure.

SUGGESTED READING

- Hill DR, ed. *Principles of diagnostic X-ray apparatus*. London/Basingstoke: Macmillan Press, 1975.
- Krestel Erich, ed. *Imaging systems for medical diagnostics*. Berlin/Munich: Siemens Aktiengesellschaft, 1990.
- McCollough CH. The AAPM/RSNA physics tutorial for residents: X-ray production. *Radiographics* 1997;17:967–984.
- Seibert JA. The AAPM/RSNA physics tutorial for residents: X-ray generators. *Radiographics* 1997;17:1533–1557.

SCREEN-FILM RADIOGRAPHY

6.1 PROJECTION RADIOGRAPHY

Projection radiography, the first radiologic imaging procedure performed, was initiated by the radiograph of the hand of Mrs. Roentgen in 1895. Radiography has been optimized and the technology has been vastly improved over the past hundred years, and consequently the image quality of today's radiograph is outstanding. Few medical devices have the diagnostic breadth of the radiographic system, where bone fracture, lung cancer, and heart disease can be evident on the same image. Although the equipment used to produce the x-ray beam is technologically mature, advancements in material science have led to improvements in image receptor performance in recent years.

Projection imaging refers to the acquisition of a two-dimensional image of the patient's three-dimensional anatomy (Fig. 6-1). Projection imaging delivers a great deal of information compression, because anatomy that spans the entire thickness of the patient is presented in one image. A single chest radiograph can reveal important diagnostic information concerning the lungs, the spine, the ribs, and the heart, because the radiographic shadows of these structures are superimposed on the image. Of course, a disadvantage is that, by using just one radiograph, the position along the trajectory of the x-ray beam of a specific radiographic shadow, such as that of a pulmonary nodule, is not known.

Radiography is a *transmission* imaging procedure. X-rays emerge from the x-ray tube, which is positioned on one side of the patient's body, they then pass through the patient and are detected on the other side of the patient by the screen-film detector. *Transmission* imaging is conceptually different from *emission* imaging, which is performed in nuclear medicine studies, or *reflection*-based imaging, such as most clinical ultrasound imaging.

In screen-film radiography, the optical density (OD, a measure of film darkening) at a specific location on the film is (ideally) determined by the x-ray attenuation characteristics of the patient's anatomy along a straight line through the patient between the x-ray source and the corresponding location on the detector (see Fig. 6-1). The x-ray tube emits a relatively uniform distribution of x-rays directed toward the patient. After the homogeneous distribution interacts with the patient's anatomy, the screen-film detector records the altered x-ray distribution.

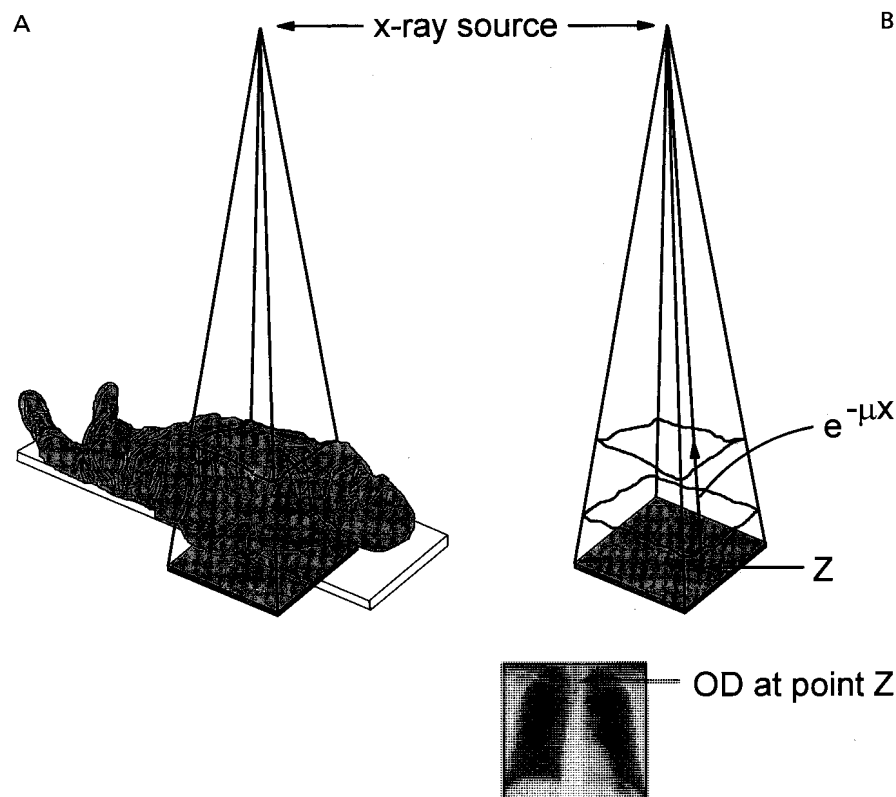


FIGURE 6-1. The basic geometry of projection radiography is illustrated. **A:** Projection radiography is a transmission imaging procedure whereby x-rays emanate from the x-ray source, pass through the patient, and then strike the detector. **B:** At each location on the film (e.g., position Z), the optical density is related to the attenuation properties ($e^{-\mu x}$) of the patient corresponding to that location.

6.2 BASIC GEOMETRIC PRINCIPLES

Two triangles that have the same shape (the three angles of one are equal to the three angles of the other) but have different sizes are said to be *similar* triangles (Fig. 6-2). If two triangles are similar, the ratios of corresponding sides and heights are equal:

$$\frac{a}{A} = \frac{b}{B} = \frac{c}{C} = \frac{h}{H} \text{ and } \frac{d}{D} = \frac{e}{E} = \frac{f}{F} = \frac{g}{G} \quad [6-1]$$

In projection radiography, similar triangles are encountered when determining image magnification and when evaluating image unsharpness caused by factors such as focal spot size and patient motion.

Magnification of the image occurs because the beam diverges from the focal spot to the image plane. For a point source, the geometry in Fig. 6-3A shows the magnification of the image size (I) in relation to the object size (O) as a function of the source-to-image distance (SID), the source-to-object distance (SOD), and the object-to-image distance (OID). The magnification (M) is defined as follows:

$$M = \frac{I}{O} \quad [6-2]$$

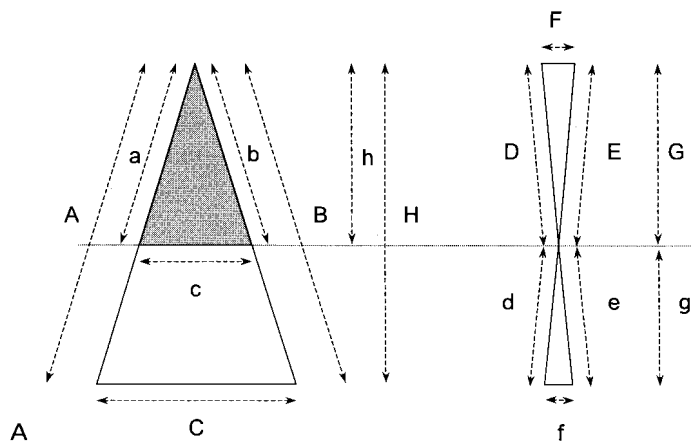


FIGURE 6-2. The sides and heights of similar triangles (those whose respective angles are equal) are proportional, according to Equation 6-1. **A:** Triangle *abc* is proportional to triangle *ABC*. The heights of triangles *abc* and *ABC* are *h* and *H*, respectively. **B:** Triangle *def* is proportional to triangle *DEF*. Triangles *def* and *DEF* have corresponding heights *g* and *G*.

where *O* is the length of the object and *I* is the corresponding length of the projected image. From similar triangles,

$$M = \frac{I}{O} = \frac{SID}{SOD} \quad [6-3]$$

The magnification is largest when the object is close to the focal spot, decreases with distance from the focal spot, and approaches a value of 1 as the object approaches the image plane.

An extended source, such as the focal spot of an x-ray tube, can be modeled as a large number of point sources projecting an image of the object on the detector plane. In Fig. 6-3B, point sources at each edge and at the center of the focal spot are shown. With magnification, geometric blurring of the object occurs in the image. Similar triangles allow the calculation of the edge gradient blurring, *f*, in terms of magnification, *M*, and focal spot size, *F*:

$$\frac{f}{F} = \frac{OID}{SOD} = \frac{SID - SOD}{SOD} = \frac{SID}{SOD} - 1 \quad [6-4]$$

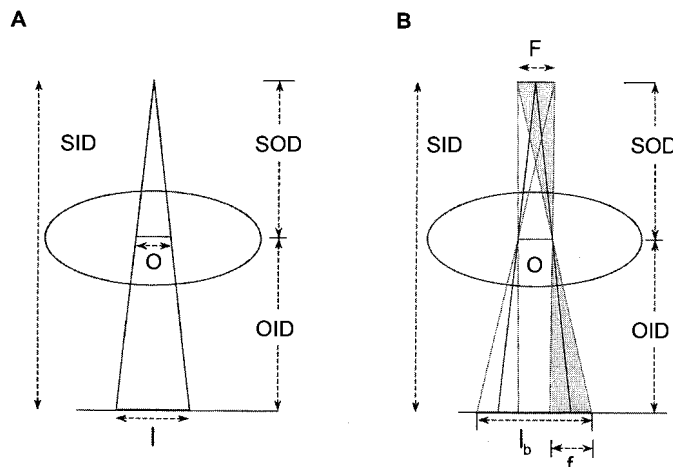


FIGURE 6-3. In projection radiography, the magnification of the object is determined by the ratio of the image, *I*, to the object, *O*. Using similar triangles, this is calculated from the source-to-image distance (*SID*), source-to-object distance (*SOD*), and object-to-image distance (*OID*). **A:** Magnification for a point source is illustrated. **B:** For a distributed source, similar triangles for one side of the focal spot and its projection are highlighted. The edge gradient increases the size of the blurred image, *I_b*, by *M* + (*M* - 1); see text for details.

Therefore,

$$f = F (M - 1)$$

Consideration of Equation 6-4 reveals that blur increases with the size of the focal spot and with the amount of magnification. Therefore, focal spot blur can be minimized by keeping the object close to the image plane.

6.3 THE SCREEN-FILM CASSETTE

The modern screen-film detector system used for general radiography consists of a *cassette*, one or two *intensifying screens*, and a sheet of *film*. The film itself is a sheet of thin plastic with a photosensitive *emulsion* coated onto one or both sides.

The cassette has several functions. Most importantly, the inside of the cassette is light tight to keep ambient light from exposing the film. After a new sheet of film has been placed in the opened cassette (in a darkroom or other dark environment), the two halves of the cassette (Fig. 6-4) are closed. During the closing process, the screens make contact with the film, at first near the cassette hinge; then, as the cassette continues to close, air is forced out of the space between the screens and the film. The screens are permanently mounted in the cassette on layers of compressible foam. The foam produces force, which holds the screens tightly against the sheet of film. Physical contact between the screens and the film is necessary to avoid artifacts on the radiographic image. By avoiding air trapping and by applying pres-

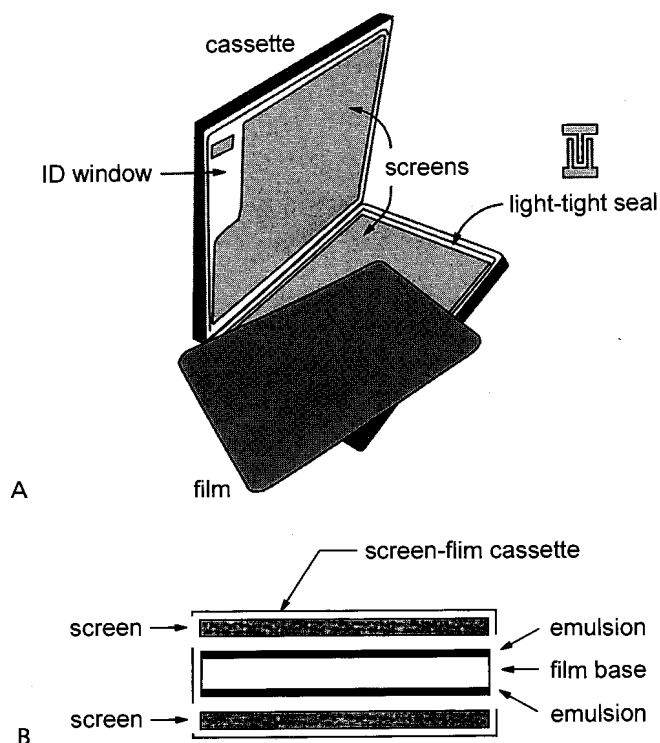


FIGURE 6-4. A: A typical screen-film cassette is illustrated in the open position. For general-purpose radiographic applications, intensifying screens are placed on the inside leaves of both the top and bottom cassette. The edges of the cassette are specially designed with a light-tight seal to avoid exposing the film to room light. **B:** A cross-section of a screen-film dual-emulsion system is illustrated.

sure on the screens, the cassette maintains good screen-film contact, which is essential for good image quality.

A screen-film cassette is designed for x-rays to enter from a particular side, usually designated with lettering (e.g., “tube side”). The presence of latches that allow the cassette to be opened indicates the back side of the cassette. The front surface of the x-ray cassette is designed to maximize the transmission of x-rays and is usually made of carbon fiber or another material of low atomic number. It is important that every image be associated with a specific patient, and a small access window on the back side of the cassette in one corner permits “flashing” of patient identification information onto the film. This information is usually provided to the technologist on an index card (*flash card*), which is placed into an *ID camera*. After the x-ray exposure, the radiologic technologist inserts the edge of the screen-film cassette into the ID camera, which opens the small cassette window in a dark chamber and optically exposes the corner of the film with an image of the flash card.

The darkroom is the location where the exposed film is removed from the cassette by hand and then placed into the chemical processor’s input chute, and a fresh sheet of film is placed into the cassette. Larger radiographic facilities make use of *daylight loading* systems, which do the work of the technologist in the darkroom automatically. In a daylight loading system, the cassette is placed into a slot and is retracted into the system (where it is dark). The exposed film is removed from the cassette and placed into the processor input chute; the system senses the size of the cassette and loads into it an appropriate sheet of fresh film; and the cassette then pops out of the system. Cassettes must be compatible with the automatic mechanisms of a daylight loading system, and special latches, which can be manipulated by mechanical actuators, are required.

6.4 CHARACTERISTICS OF SCREENS

Film by itself can be used to detect x-rays, but it is relatively insensitive and therefore a lot of x-ray energy is required to produce a properly exposed x-ray film. To reduce the radiation dose to the patient, x-ray screens are used in all modern medical diagnostic radiography. Screens are made of a scintillating material, which is also called a *phosphor*. When x-rays interact in the phosphor, visible or ultraviolet (UV) light is emitted. In this sense, the intensifying screen is a radiographic transducer—it converts x-ray energy into light. It is the light given off by the screens that principally causes the film to be darkened; only about 5% of the darkening of the film is a result of direct x-ray interaction with the film emulsion. Screen-film detectors are considered an *indirect* detector, because the x-ray energy is first absorbed in the screen or screens, and then this pattern is conveyed to the film emulsion by the visible or UV light—a two-step, indirect process.

Composition and Construction

For most of the 20th century, calcium tungstate (CaWO_4) was the most common scintillator used in x-ray intensifying screens. Thomas Edison first identified the excellent imaging properties of this phosphor in the early 1900s. In the early 1970s, rare earth phosphors were introduced that outperformed and eventually replaced CaWO_4 in intensifying screens worldwide. The term *rare earth* refers to the rare

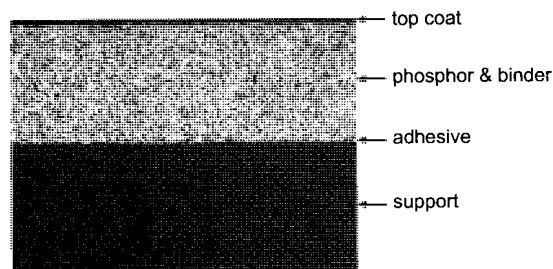


FIGURE 6-5. A scanning electron micrograph of an intensifying screen is pictured. The layers have been labelled. SEM image courtesy of Drs. Bernard Apple and John Sabol.

earth row of the periodic table, which includes elements 57 through 71. The most common rare earth phosphor used in today's intensifying screens is gadolinium oxysulfide ($\text{Gd}_2\text{O}_2\text{S}$); however lanthanum oxybromide (LaOBr), yttrium tantalate (YTaO_4) and other phosphors are used as well. Cesium iodide (CsI) is an excellent scintillator and is used in fluoroscopy and digital radiography; however, it is too moisture sensitive and fragile to be used in screen-film radiography.

A cross-sectional image of an intensifying screen is shown in Fig. 6-5. To produce an intensifying screen, the x-ray phosphor is mixed at high temperature with a small quantity of screen binder (a polymer that holds the phosphor particles together), and while it is molten the mixture is evenly spread onto a very thin layer of plastic. After hardening, a thick plastic support layer is glued onto the top of the x-ray phosphor. The thick glued-on plastic layer is the support shown in Fig. 6-5, and the thin plastic layer is the topcoat. The topcoat serves to protect the scintillator layer in a screen-film cassette from mechanical wear, which results from the frequent exchange of film. The optical properties at the interface between the phosphor layer and the top coat are most important because light must pass through this interface to reach the film emulsion, whereas the optical properties at the interface between the phosphor layer and the screen support are less important because light reaching this interface rarely reaches the film emulsion. This is why the screen phosphor is laid directly onto the topcoat during manufacture.

The phosphor thickness is usually expressed as the mass thickness, the product of the actual thickness (cm) and the density (g/cm^3) of the phosphor (excluding the binder). For general radiography, two screens are used, with each screen having a phosphor thickness of about $60 \text{ mg}/\text{cm}^2$, for a total screen thickness of $120 \text{ mg}/\text{cm}^2$. For the high-resolution requirements and low energies used in mammography, a single screen of approximately $35 \text{ mg}/\text{cm}^2$ is used.

Intensifying Screen Function and Geometry

The function of an intensifying screen is two-fold: it is responsible for absorbing (detecting) incident x-rays, and it emits visible (or UV) light, which exposes the film emulsion. When an x-ray photon interacts in the scintillator and deposits its energy, a certain fraction of this energy is converted to visible light. The conversion efficiency of a phosphor is defined as the fraction of the absorbed energy that is emitted as light. For the older CaWO_4 phosphor, about 5% of the detected x-ray energy is converted to light energy. $\text{Gd}_2\text{O}_2\text{S}:\text{Tb}$ has a higher *intrinsic* conversion efficiency of about 15%, and that is one reason why it has replaced CaWO_4 in most modern radiology departments. The light emitted from $\text{Gd}_2\text{O}_2\text{S}:\text{Tb}$ is green and has an average wavelength of 545 nm, corresponding to an energy of 2.7 eV per photon. With a 15% intrinsic con-

version efficiency, the approximate number of green light photons produced by the absorption of a 50-keV x-ray photon would be calculated as follows:

$$50,000 \text{ eV} \times 0.15 = 7,500 \text{ eV}$$

$$7,500 \text{ eV} / 2.7 \text{ eV/photon} \approx 2,800 \text{ photons}$$

Although 2,800 photons may be produced at the site of the x-ray interaction, the number that actually reaches the film emulsion after diffusing through the phosphor layer and being reflected at the interface layers is in the range of 200 to 1,000 photons.

The x-ray screen should detect as many incident x-ray photons as is physically possible: x-rays that pass through both intensifying screens go undetected and are essentially wasted. The *quantum detection efficiency* (QDE) of a screen is defined as the fraction of incident x-ray photons that interact with it. The easiest way to increase the screen's QDE is to make it thicker (Fig. 6-6). However, after an x-ray absorption event, the visible light given off in the depth of a screen propagates toward the screen's surface to reach the adjacent emulsion layer. As light propagates through the screen, it spreads out in all directions with equal probability (isotropic diffusion). For thicker screens, consequently, the light photons propagate greater lateral distances before reaching the screen surface (Fig. 6-7). This lateral diffusion of light, which increases with thicker screens, causes a slightly blurred image (Fig. 6-8). For the thicker general radiographic x-ray screens, a sandwiched dual-screen dual-emulsion screen-film combination is used (see Fig. 6-7A, B). In this approach the total screen thickness is substantial, but by placing the dual-emulsion film in the middle of the two screens the average light diffusion path is reduced, and this geometry reduces blurring. When higher spatial resolution is necessary, a single, thin screen is used with a single-emulsion film (Fig. 6-7C); this configuration is used in mammography and in some *detail* cassettes. For a single-screen single-emulsion system, the screen is placed on the opposite side of the film from the patient and the x-ray source, and of course the film emulsion faces the screen. This orientation is used because the majority of x-rays that are incident on the screen interact near the top surface, and fewer x-rays interact toward the bottom surface. The geometry

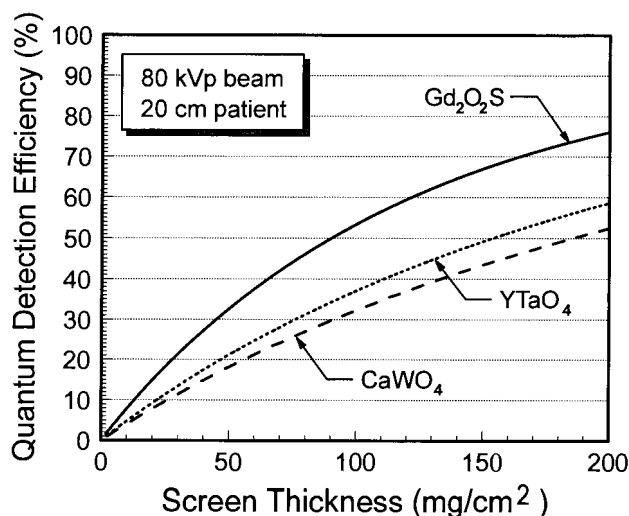


FIGURE 6-6. Quantum detection efficiency (QDE) as a function of screen thickness is illustrated for three common intensifying screen phosphors. An 80-kVp beam filtered by a 20-cm-thick patient was used for the calculation. The QDE corresponds to the fraction of x-ray photons attenuated by the intensifying screens.

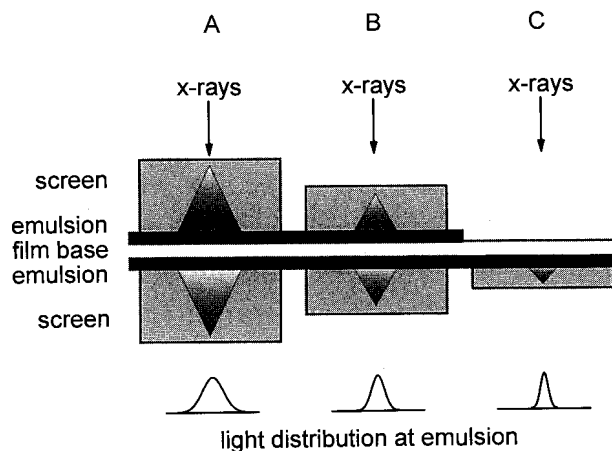


FIGURE 6-7. A cross-section of a screen-film cassette with x-ray interaction events is shown. **A:** X-rays incident on and absorbed by a thick screen-film detector produce light photons in the screen matrix that propagate a great distance. The thicker intensifying screens absorb a greater fraction of the incident x-rays, but a greater lateral spread of the visible light occurs, with a consequent reduction in spatial resolution. **B:** A thin radiographic screen-film detector results in less x-ray absorption, but less lateral spread of visible light and better spatial resolution is achieved. **C:** For x-ray imaging procedures requiring maximum spatial resolution, a single-screen single-emulsion radiographic cassette is used. For such systems, the x-rays first traverse the film emulsion and base and then strike the intensifying screen. Because most of the interactions occur near the top of the intensifying screen, this geometry results in less light spread and better spatial resolution.

shown in Fig. 6-7C therefore results in a shorter diffusion path and higher resolution than if the screen were placed on top of the film. The geometry in Fig. 6-7C is possible only because x-rays easily penetrate the film.

For dual-phosphor dual-emulsion systems, it is very important that the light from the top screen not penetrate the film base and expose the bottom emulsion

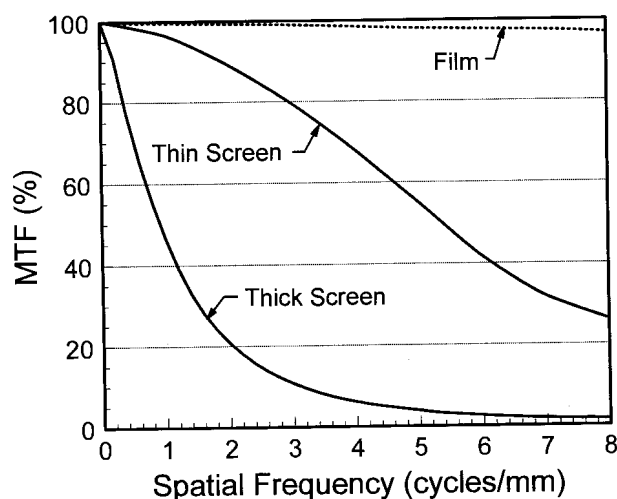


FIGURE 6-8. The modulation transfer function (MTF) as a function of spatial frequency is illustrated for intensifying screens of different thicknesses. The thicker screen corresponds to a 120 mg/cm² thickness of Gd₂O₂S, and the thinner screen is for a 34 mg/cm² thickness of Gd₂O₂S. Thicker screens have better absorption properties but exhibit poorer spatial resolution. The MTF of film only is illustrated, and film by itself has far better spatial resolution than that of intensifying screens. In a screen-film cassette, the resolution properties of the screen determine the overall spatial resolution.

(and vice versa). This phenomenon is called *print-through* or *crossover*. This was more of a problem in the past, when spherical-grained films were used for screen-film radiography. With the development of so-called T-grain film emulsions (flat-shaped grain geometry), crossover has become much less of a problem.

It is desirable to have the intensifying screen absorb as many of the x-ray photons incident upon it as possible. Although increasing the screen's thickness increases detection efficiency and thereby improves the screen's *sensitivity*, increased thickness also causes an undesirable loss of *spatial resolution*. This phenomenon illustrates the classic compromise between sensitivity and resolution seen with many imaging systems. The *modulation transfer function* (MTF) is the technical description of spatial resolution for most imaging systems. MTFs are shown in Fig. 6-8 for two screen-film systems having different screen thicknesses. The thinner screen has a much better MTF and therefore better spatial resolution performance; however, the thinner screen has significantly less QDE at diagnostic x-ray energies. MTFs are discussed in more detail in Chapter 10.

Conversion Efficiency

It was mentioned earlier that $\text{Gd}_2\text{O}_2\text{S:Tb}$ and CaWO_4 have *intrinsic* conversion efficiencies of about 15% and 5%, respectively. The total conversion efficiency of a screen-film combination refers to the ability of the screen or screens to convert the energy deposited by *absorbed* x-ray photons into film darkening. The overall conversion efficiency depends on the intrinsic conversion efficiency of the phosphor, the efficiency of light propagation through the screen to the emulsion layer, and the efficiency of the film emulsion in absorbing the emitted light. The light propagation in the screen is affected by the transparency of the phosphor to the wavelength of light emitted. Because light propagation in the lateral direction causes a loss in spatial resolution, a *light-absorbing dye* is added to some screens to reduce the light propagation distance and preserve spatial resolution. The presence of such a dye reduces the conversion efficiency of the screen-film system (Fig. 6-9). A variation of

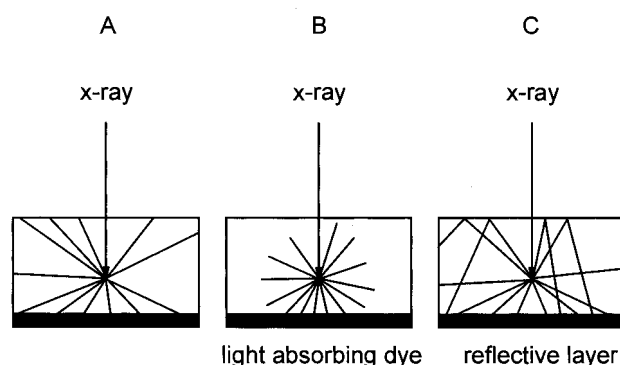


FIGURE 6-9. A: The light in a phosphor screen diffuses in all directions and causes a loss of spatial resolution. **B:** Light-absorbing dyes in the phosphor material reduce the average distance that light photons travel and increase the spatial resolution, but they also reduce light conversion efficiency. **C:** Screen-film cassettes may employ a reflective layer placed on the back of the screen surface to redirect light photons toward the film emulsion and increase conversion efficiency. However, this causes greater lateral spreading and reduces spatial resolution.

the use of dye is to design a phosphor that emits in the UV region of the optical spectrum, because UV light is more readily absorbed in the phosphor layer than are the longer wavelengths of light. To produce a faster (i.e., higher conversion efficiency), lower-resolution screen-film system, a *reflective layer* can be placed between the screen and its support. By this means, light that is headed away from the film emulsion is redirected back toward the emulsion.

An intensifying screen by itself is a linear device at a given x-ray energy. If the number of x-ray photons is doubled, the light intensity produced by the screen also doubles, and a tenfold increase in x-rays produces a tenfold increase in light. This input-output relationship for the intensifying screen remains linear over the x-ray exposure range found in diagnostic radiology.

Absorption Efficiency

The absorption efficiency, or more precisely the QDE, describes how efficiently the screen detects x-ray photons that are incident upon it. The QDEs of three different intensifying screens are shown in Fig. 6-10 as a function of the peak kilovoltage (kVp) of the x-ray beam. As an x-ray *quantum* (a photon) is absorbed by the screen, the energy of the x-ray photon is deposited and some fraction of that energy is converted to light photons. However, the x-ray beam that is incident upon the detector in radiography is polyenergetic and therefore has a broad spectrum of x-ray energies. For example, a 100-kVp beam contains x-rays from about 15 to 100 keV. When an 80-keV x-ray photon is absorbed in the screen, it deposits twice the energy of a 40-keV x-ray. The number of light photons produced in the intensifying screen (which is related to film darkening) is determined by the total amount of x-ray *energy* absorbed by the screen, not by the number of x-ray *photons*. Therefore, screen-film systems are considered *energy detectors*. This contrasts with the function of detectors used in nuclear medicine applications, which count gamma ray photons individually and are considered *photon counters*.

The $\text{Gd}_2\text{O}_2\text{S}$ phosphor demonstrates an impressive improvement in QDE over the other phosphors. However, when the QDE (see Fig. 6-10) is compared with the

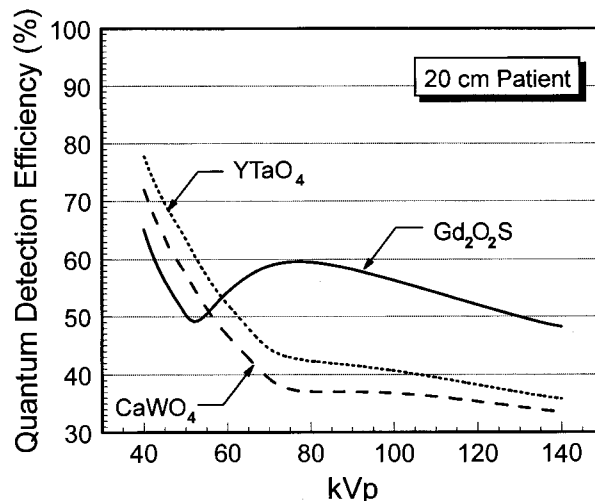


FIGURE 6-10. Quantum detection efficiency (QDE) is shown as a function of the peak kilovoltage (kVp) of the x-ray beam for three common intensifying screen phosphors for a 120 mg/cm² screen thickness. The x-ray spectra are first passed through a 20-cm-thick water-equivalent patient to mimic patient attenuation. A notable increase in QDE of the $\text{Gd}_2\text{O}_2\text{S}$ phosphor is observed at energies higher than 50 keV (the k-edge of Gd).

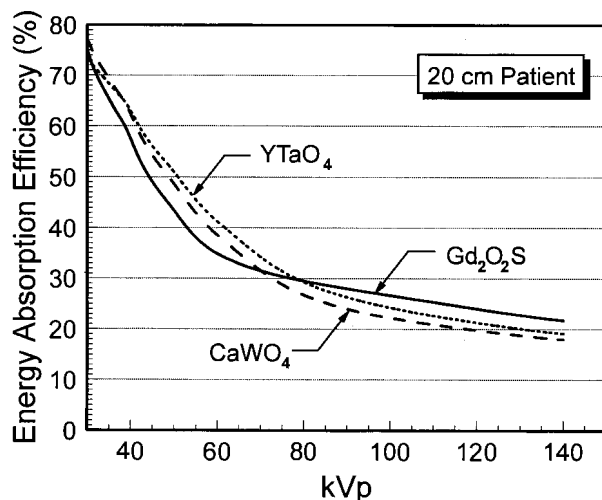


FIGURE 6-11. The energy absorption efficiency is shown as a function of the peak kilovoltage (kVp) of the x-ray beam for three intensifying screen phosphors. The x-ray beam is passed through 20 cm of patient, and an intensifying screen thickness of 120 mg/cm² is used. The escape of characteristic x-ray photons in Gd₂O₂S partially offsets the increased quantum detection efficiency shown in Fig. 6-10.

energy absorption efficiency (Fig. 6-11), the improvement is seen to be more modest. The increase in QDE for Gd₂O₂S at energies greater than 50 keV is caused by the large increase in photon absorption at and above the k-edge of Gd (50.3 keV). As the kVp increases from 55 to 75 kVp, for example, there is a shift in the x-ray spectrum whereby an increasing number of x-ray photons are greater than 50.3 keV. The increase in kVp in this voltage range results in better x-ray photon absorption. However, photoelectric absorption usually results in the re-emission of a characteristic x-ray photon. Discrete characteristic x-ray energies for Gd occur between 42 to 50 keV. After an incident x-ray photon is captured in the screen by photoelectric absorption, the characteristic x-rays are re-emitted from the screen and their energy can be lost. Consequently, although the initial photon detection improves above the k-edge, the ultimate energy absorption efficiency does not enjoy a similar advantage.

Overall Efficiency of a Screen-Film System

The spatial resolution of x-ray film in the absence of a screen is almost perfect, so given the detrimental role that screens play on spatial resolution (see Fig. 6-8), why are they used? Screens are used in medical radiography for one reason: to reduce the radiation dose to the patient. The total efficiency of a screen-film system is the product of the conversion efficiency and the absorption efficiency. A screen-film system provides a vast increase in x-ray detection efficiency over film by itself. For example, at 80 kVp, a 120 mg/cm² thickness of Gd₂O₂S detects 29.5% of the incident x-ray energy, whereas film detects only 0.65% of the incident x-ray energy. Figure 6-12 illustrates the improvement of screen-film systems over film only. For general diagnostic radiographic procedures in the 80- to 100-kVp range, use of the intensifying screen improves detection efficiency by approximately a factor of 50. This means that patient dose is reduced by a factor of 50, a substantial amount. There are other tangible benefits of using screens. For example, the output requirements of the x-ray system are reduced, substantially reducing the requirement for powerful x-ray generators and high heat capacity x-ray tubes and reducing costs.

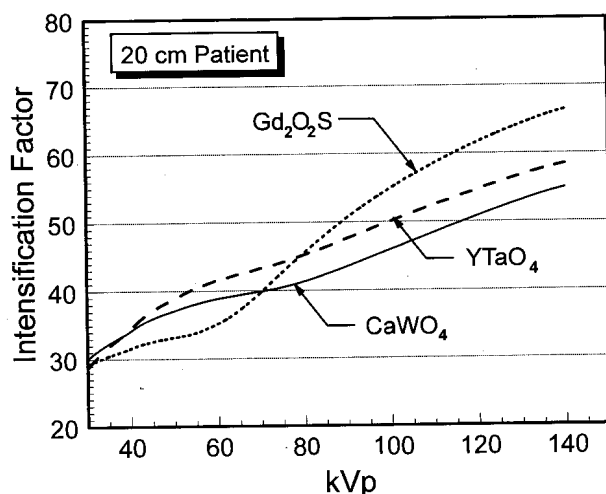


FIGURE 6-12. The intensification factor is illustrated as a function of the peak kilovoltage (kVp) of the x-ray beam for three intensifying screen phosphors. The intensification factor is the ratio of the energy absorption of the 120 mg/cm² phosphor to the energy absorption of 0.80 mg/cm² of AgBr. An intensification factor of approximately 50 is achieved for Gd₂O₂S over the x-ray energy spectra commonly used in diagnostic examinations.

Because the x-ray levels required for proper exposure are reduced, the exposure times are shorter and motion artifacts are reduced. The exposure to radiology personnel from scattered x-rays is reduced as well.

Noise Effects of Changing Conversion Efficiency versus Absorption Efficiency

The term *noise* refers to local variations in film OD that do not represent variations in attenuation in the patient. Noise includes *random noise*, caused by factors such as random variations in the numbers of x-ray photons interacting with the screen, random variations in the fraction of light emitted by the screen that is absorbed in the film emulsion, and random variations in the distribution of silver halide grains in the film emulsion. As discussed in detail in Chapter 10, the noise in the radiographic image is governed principally by the number of x-ray photons that are detected in the screen-film detector system. The visual perception of noise is reduced (resulting in better image quality) when the number of *detected* x-ray photons increases.

What happens to the noise in the image when the conversion efficiency of a screen-film system is increased, such as by adding a reflective layer to the screens? To answer, it must be realized that a properly exposed film is required. Consequently, if the “speed” of the screen-film system is increased by increasing the conversion efficiency (so that each detected x-ray photon becomes more efficient at darkening the film), fewer detected x-ray photons are required to achieve the same film darkening. Fewer detected x-rays result in more radiographic noise. Therefore, increasing the conversion efficiency to increase the speed of a screen-film system will increase the noise in the images.

What happens to the noise in the image when the absorption efficiency is increased, such as by making the screen thicker? As before, the number of detected x-ray photons before and after the proposed change must be compared at the same

film OD. If the screen is made thicker so that 10% more x-ray photons are detected with no change in the conversion efficiency, then a reduction in the incident x-ray beam of 10% is required to deliver the same degree of film darkening. The fraction of x-ray photons that are detected increases, but the number of incident x-ray photons must be reduced by that same fraction, so the total number of detected x-ray photons remains the same. Consequently, the noise in the image is not affected by changing the absorption efficiency. However, spatial resolution suffers when the screen is made thicker.

6.5 CHARACTERISTICS OF FILM

Composition and Function

Unexposed film consists of one or two layers of film emulsion coated onto a flexible sheet made of Mylar (a type of plastic). Tabular grain emulsions (Fig. 6-13A) are used in modern radiographic film; laser cameras and older radiographic films use cubic grain film (see Fig. 6-13B). The grains of silver halide (AgBr and AgI) are bound in a gelatin base and together comprise the film emulsion. A typical sheet of dual-

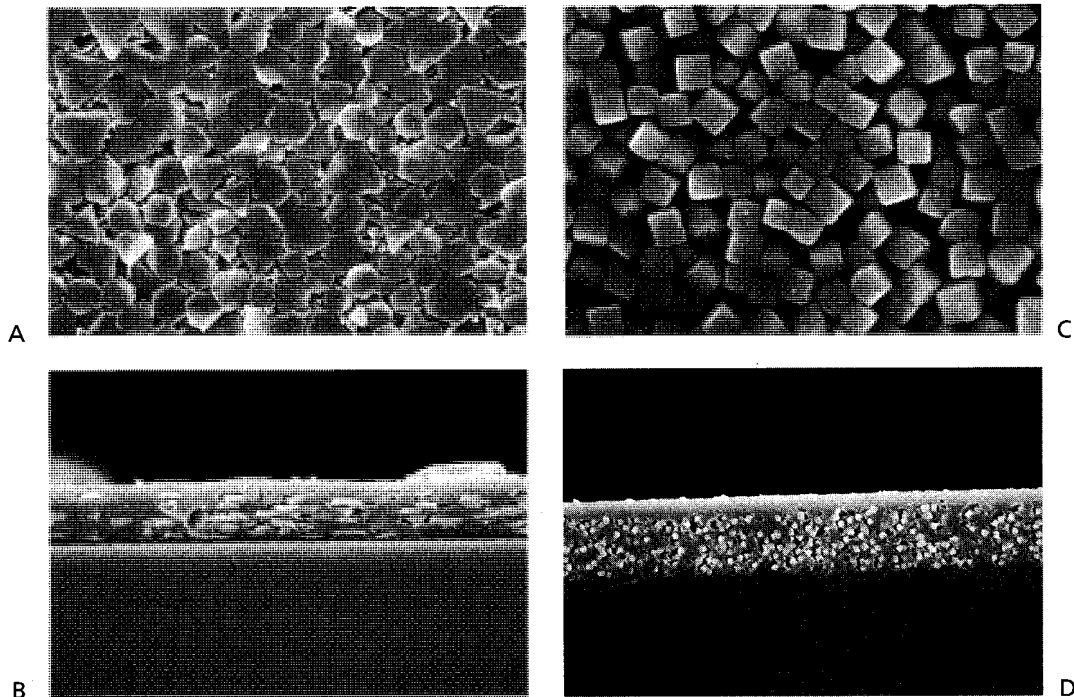


FIGURE 6-13. Scanning electron micrographs (SEM) are shown. **A:** A top-down SEM image of the emulsion layer of T grain emulsion is shown. **B:** A cross section of the T grain emulsion film is illustrated, showing the grains in the gelatin layer, supported by the polymer film base below. **C:** Cubic grain film emulsion is shown in a top-down SEM view. **D:** A cross sectional SEM image of the cubic grain film is illustrated. SEM photographs courtesy of Drs. Bernard Apple and John Sabol.

emulsion film has a total mass thickness of 0.80 mg/cm^2 of silver halide, and single-emulsion films have about 0.60 to 0.70 mg/cm^2 of silver halide in their single emulsion layer.

The emulsion of an exposed sheet of x-ray film contains the *latent* image. The exposed emulsion does not look any different than an unexposed emulsion, but it has been altered by the exposure to light—the latent image is recorded as altered chemical bonds in the emulsion, which are not visible. The latent image is rendered visible during film processing by chemical reduction of the silver halide into metallic silver grains. A film processor performs chemical processing, the details of which are discussed in Chapter 7.

Optical Density

X-ray film is a negative recorder, which means that increased light exposure (or x-ray exposure) causes the developed film to become darker. This is the opposite of the case with photographic print film, in which more exposure produces brighter response. The degree of darkness of the film is quantified by the OD, which is measured with a *densitometer*. The densitometer is a simple device that shines white light onto one side of the developed film and electronically measures the amount of light reaching the other side. The densitometer has a small sensitive area (aperture), typically about 3 mm in diameter, and measurements of OD correspond to that specific area of the film. If the intensity of the light measured with no film in the densitometer is given by I_0 , and the intensity measured at a particular location on a film is given by I , then the *transmittance* (T) of the film at that location and the OD are defined as follows:

$$T = \frac{I}{I_0}$$

and the OD is:

$$\text{OD} = -\log_{10}(T) = \log_{10}\left(\frac{1}{T}\right) = \log_{10}\left(\frac{I_0}{I}\right)$$

If the transmission through the film is $T = 0.1 = 10^{-1}$, then $\text{OD} = 1$; if the transmission is $T = 0.01 = 10^{-2}$, then $\text{OD} = 2$. The relationship between OD and T is analogous to that between pH and the hydrogen ion concentration ($\text{pH} = -\log_{10} [\text{H}_3\text{O}^+]$). The inverse relationship is $T = 10^{-\text{OD}}$. Table 6-1 lists examples of ODs and corresponding transmission values for diagnostic radiology.

TABLE 6-1. RELATIONSHIP BETWEEN OPTICAL DENSITY (OD) AND TRANSMISSION (T) PERTINENT TO DIAGNOSTIC RADIOLOGY APPLICATIONS

T	T	OD	Comment
1.0000	10^{-0}	0	Perfectly clear film (does not exist)
0.7760	10^{-11}	0.11	Unexposed film (base + fog)
0.1000	$10^{-1.0}$	1	Medium gray
0.0100	$10^{-2.0}$	2	Dark
0.0010	$10^{-3.0}$	3	Very dark; requires hot lamp
0.00025	$10^{-3.6}$	3.6	Maximum OD used in medical radiography

The Hurter and Driffield Curve

Film has excellent spatial resolution, far beyond that of the screen (see Fig. 6-8). Film used for microfiche, for example, is capable of spatial resolutions approaching 100 cycles per millimeter. The more interesting property of film in the context of screen-film radiography, therefore, is how film responds to x-ray or light exposure.

If a sheet of fresh unexposed film is taken out of the box and processed, it will have an OD in the range from about 0.11 to 0.15. This OD is that of the *film base*. The polyester film base is usually slightly tinted to hide faint chemical residue from the processing, which can appear as an unattractive brown. Although gray-tinted film base is available, a blue tint is considered by many to be pleasing to the eye and is used quite commonly. Film that has been stored for a long period or exposed to heat or to background radiation can develop a uniform fog level. In this case, an otherwise unexposed sheet of film after processing may have an OD in the 0.13 to 0.18 range. The OD resulting from the combination of background fogging and the tinting of the film base is referred to as the “base + fog” OD. Base + fog levels exceeding about 0.20 are considered unacceptable, and replacement of such film should be considered.

When film is exposed to the light emitted from an intensifying screen in a screen-film cassette, its response as a function of x-ray exposure is nonlinear (Fig. 6-14). Hurter and Driffield studied the response of film to light in the 1890s, and the graph describing OD versus the logarithm (base 10) of exposure is called the *H & D curve*, in their honor. A more generic term is the *characteristic curve*. Notice that the x-axis of the H & D curve (see Fig. 6-14) is on a logarithmic scale, and this axis is often called the *log relative exposure*. The OD (the value on the y-axis) is itself the logarithm of the transmission, and therefore the H & D curve is a $\log_{10}$ - $\log_{10}$ plot of *optical transmission* versus *x-ray exposure*.

Salient features of the H & D curve are shown in Fig. 6-14. The curve has a sigmoid shape. The *toe* is the low-exposure region of the curve. Areas of low exposure on a radiograph are said to be “in the toe” of the H & D curve. For example, the mediastinum on a chest radiograph is almost always in the toe. The toe region extends down to zero exposure (which cannot actually be plotted on a logarithmic

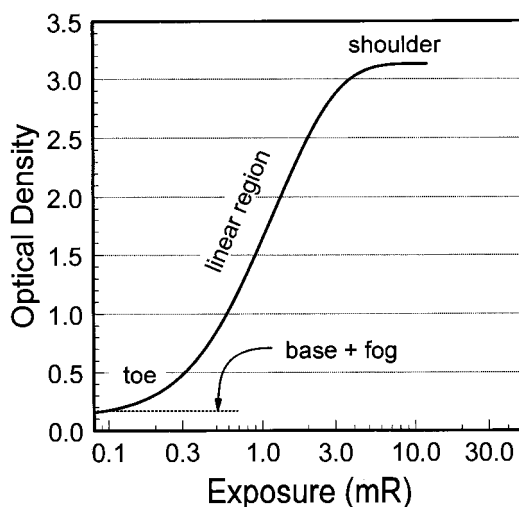


FIGURE 6-14. A Hurter and Driffield (H & D) curve is a plot of a film's optical density (OD) as a function of the log of exposure. The regions of the H & D curve include the toe, the linear region, and the shoulder. The base + fog level corresponds to the OD of unexposed film.

axis), and the film OD at zero x-ray exposure corresponds to the *base + fog* level of the film. Beyond the toe is the *linear region* of the H & D curve. Ideally, most of a radiographic image should be exposed in this region. The high-exposure region of the H & D curve is called the *shoulder*.

The contrast of a radiographic film is related to the slope of the H & D curve: regions of higher slope have higher contrast, and regions of reduced slope (i.e., the toe and the shoulder) have lower contrast. A single number, which defines the overall contrast level of a given type of radiographic film, is the *average gradient* (Fig. 6-15). The average gradient is the slope of a straight line connecting two well-defined points on the H & D curve. The lower point is usually defined at $OD_1 = 0.25 + \text{base} + \text{fog}$, and the higher point is typically defined at $OD_2 = 2.0 + \text{base} + \text{fog}$. To calculate the average gradient, these two OD values are identified on the y-axis of the H & D curve, and the corresponding exposures, E_1 and E_2 , are then identified. The average gradient is the slope ("the rise over the run") of the curve:

$$\text{Average gradient} = \frac{OD_2 - OD_1}{\log_{10}(E_2) - \log_{10}(E_1)}$$

Average gradients for radiographic film range from 2.5 to 3.5. The average gradient is a useful parameter that describes the contrast properties of a particular film, but the H & D curve is very nonlinear and therefore the slope of the curve (the contrast) actually depends on the exposure level. Fig. 6-16A illustrates the slope of the H & D curve (see Fig. 6-15), plotted as a function of the logarithm of x-ray exposure. The contrast-versus-exposure graph demonstrates how important it is to obtain the correct exposure levels during the radiographic examination. Because high contrast is desirable in radiographic images, optimal exposures occur in the region near the maximum of the contrast curve. If the exposure levels are too high or too low, contrast will suffer. Figure 6-16B illustrates the contrast as a function of OD. Because OD is something that can easily be measured on the film after the exposure and film processing, the graph shown in Fig. 6-16B is useful in determin-

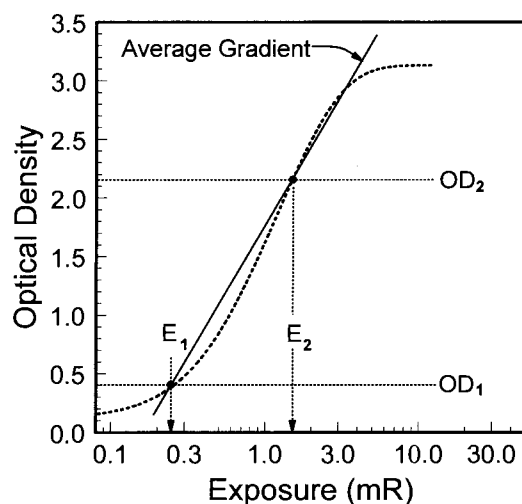


FIGURE 6-15. The average gradient is measured between two specific optical densities on the Hurter and Driffield (H & D) curve: $OD_1 = 0.25 + \text{base} + \text{fog}$, and $OD_2 = 2.0 + \text{base} + \text{fog}$. The corresponding exposure levels (E_1 and E_2 , respectively) are determined from the H & D curve, and the average gradient is calculated as the slope between these two points on the curve.

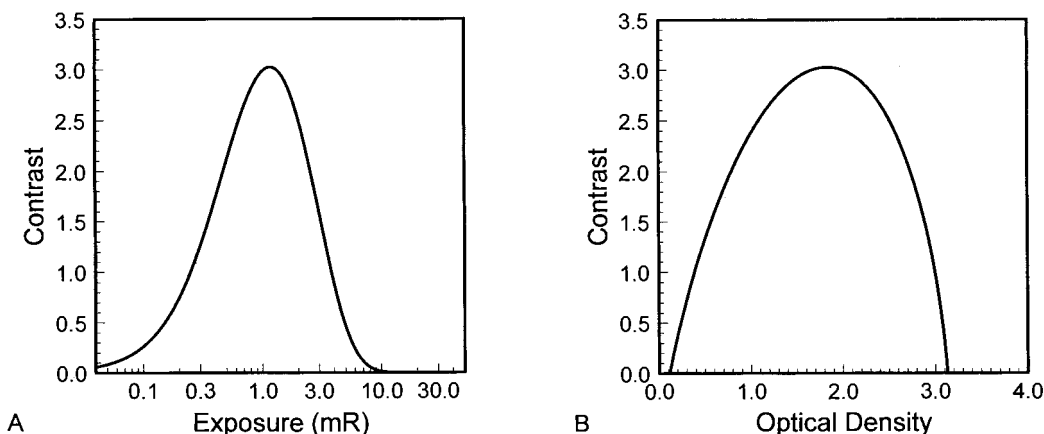


FIGURE 6-16. A: The contrast of the screen-film system shown in Fig. 6-15 is illustrated as a function of exposure, by computing (by differentiation) the slope at every point along the Hurter and Driffield (H & D) curve. Because the H & D curve is nonlinear, the slope changes along the curve. **B:** The contrast as a function of optical density (OD) is derived from the H & D curve shown in Fig. 6-15. Maximum contrast in the OD range from approximately 1.5 to 2.5 is achieved for this screen-film system.

ing what OD range should be achieved for a given radiographic screen-film system. The film manufacturer physically controls the contrast on a film by varying the size distribution of the silver grains. High-contrast films make use of a homogeneous distribution of silver halide grain sizes, whereas lower-contrast films use a more heterogeneous distribution of grain sizes.

The sensitivity or *speed* of a screen-film combination is evident from the H & D curve. As the speed of a screen-film system increases, the amount of x-ray exposure required to achieve the same OD decreases. Faster (higher-speed) screen-film systems result in lower patient doses but in general exhibit more quantum mottle than slower systems. There are two different measures of the speed of a screen-film combination. The *absolute speed* is a quantitative value that can be determined from the H & D curve. The absolute speed of a screen-film combination is simply the inverse of the exposure (measured in roentgens) required to achieve an $OD = 1.0 + \text{base} + \text{fog}$. Figure 6-17 shows two H & D curves, for screen-film systems A and B. System A is faster than system B, because less radiation is required to achieve the same OD. The exposures corresponding to the $OD = 1.0 + \text{base} + \text{fog}$ levels for these systems are approximately 0.6 and 2.0 mR, corresponding to absolute speeds of $1,667 \text{ R}^{-1}$ ($1/0.0006 \text{ R}$) and 500 R^{-1} ($1/0.002 \text{ R}$), respectively. Absolute speed is usually used in scientific discussions concerning the performance of screen-film systems.

The commercial system for defining speed makes use of a relative measure. When CaWO_4 screens were in common use, so-called *Par* speed systems were arbitrarily given a speed rating of 100. The speed of other screen-film systems in a vendor's product line are related to the Par system for that vendor—so, for example, a 200-speed system is about twice as fast as a 100-speed system. Today, with rare earth screen-film combinations prevailing, most institutions use 400-speed systems for general radiography. Curve A shown in Fig. 6-17 was measured on a 400-speed system. Slower-speed screen-film combinations (i.e., 100 speed) are often used for

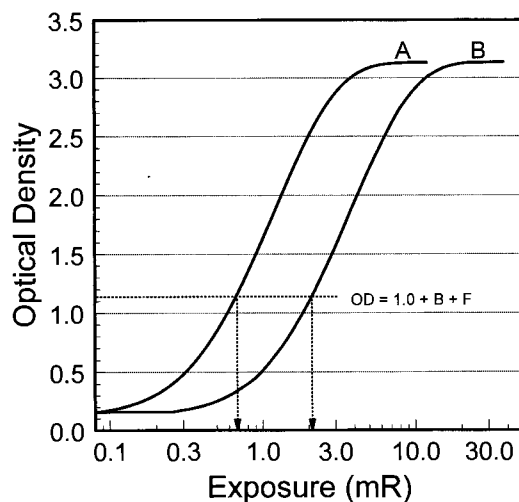


FIGURE 6-17. Hurter and Driffield (H & D) curves are shown for two different screen-film detectors, system A and system B. A lateral shift in H & D curve indicates a change in the speed (or sensitivity) between systems. The horizontal dotted line corresponding to $OD = 1.0 + \text{base} + \text{fog}$ indicates the point at which the speed is evaluated. The x-ray exposures required to reach this OD level are indicated for each system (vertical arrows). System A requires less exposure than system B to achieve the same OD; therefore, it is the faster, more sensitive of the two screen-film systems.

detail work, typically bone radiographs of the extremities. Screen-film systems in the 600-speed class are used in some radiology departments for specialty applications (e.g., angiography) where short exposure times are very important. To put the sensitivity of typical screen-film systems in more concrete terms, a 400-speed, 35- $\times$ 43-cm cassette (requiring about 0.6 mR to achieve an OD of 1.0 + base + fog), producing a uniform gray of about 1.15 OD, is exposed to about 2.4×10^{10} (i.e., 24 billion) x-ray photons, corresponding to about 156,000 x-ray photons per square millimeter.

Whereas a horizontal shift between two H & D curves (see Fig. 6-17) demonstrates that the systems differ in speed, systems with different contrast have H & D curves with different slopes (Fig. 6-18). Contrast is desirable in screen-film radiography, but compromises exist. Screen-film system A in Fig. 6-18 has higher contrast than system B, because it has a steeper slope. A disadvantage of high contrast is

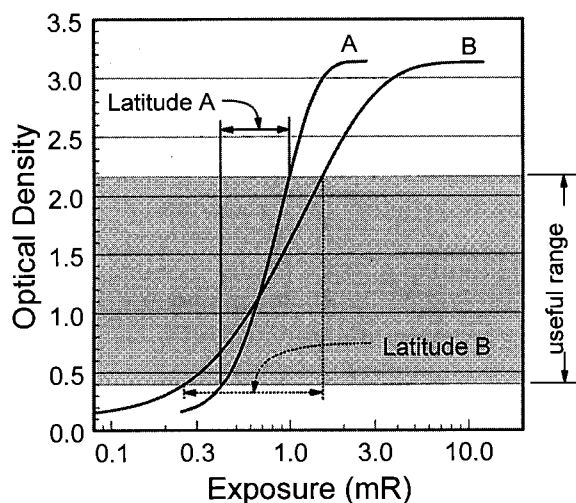


FIGURE 6-18. A Hurter and Driffield (H & D) curve illustrates screen-film systems A and B, which differ in their contrast. The slope of system A is steeper than that of system B, and therefore system A has higher contrast. The latitude of a screen-film system refers to the range of exposures that produce acceptable optical densities. The compromise of a high-contrast screen-film system (i.e., system A) is that it has reduced latitude.

reduced *latitude*. The shaded region on Fig. 6-18 illustrates the useable range of optical densities, and the latitude is the range of x-ray exposures that deliver ODs in the usable range. Latitude is called *dynamic range* in the engineering world. System A in Fig. 6-18 has higher contrast but lower latitude. It is more difficult to consistently achieve proper exposures with low-latitude screen-film systems, and therefore these systems contribute to higher retake rates. In the case of chest radiography, it may be impossible to achieve adequate contrast of both the mediastinum and the lung fields with a low-latitude system.

6.6 THE SCREEN-FILM SYSTEM

The film emulsion should be sensitive to the wavelengths of light emitted by the screen. CaWO_4 emits blue light to which silver halide is sensitive. Native silver halide is not as sensitive to the green or red part of the visible spectrum. Consequently, in the days before the introduction of rare earth screens, darkrooms had bright red or yellow safelights that had little effect on the films of the day. The emission spectrum of $\text{Gd}_2\text{O}_2\text{S:Tb}$ is green, and to increase the sensitivity of silver halide to this wavelength sensitizers are added to the film emulsion. Modern film emulsions that are green-sensitized are called *orthochromatic*, and film emulsions that are sensitized all the way into the red region of the visible spectrum (e.g., color slide film) are called *panchromatic*. Because matching the spectral sensitivity of the film to the spectral output of the screen is so important, and for other compatibility reasons, the screens and films are usually purchased as a combination—the *screen-film combination* or *screen-film system*. Combining screens from one source and film from another is not a recommended practice unless the film manufacturer specifies that the film is designed for use with that screen.

The reciprocity law of film states that the relationship between exposure and OD should remain constant regardless of the exposure rate. This is equivalent to saying that the H & D curve is constant as a function of exposure rate. However, for very long or very short exposures, an exposure *rate* dependency between exposure and OD is observed, and this is called *reciprocity law failure* (Fig. 6-19). At long

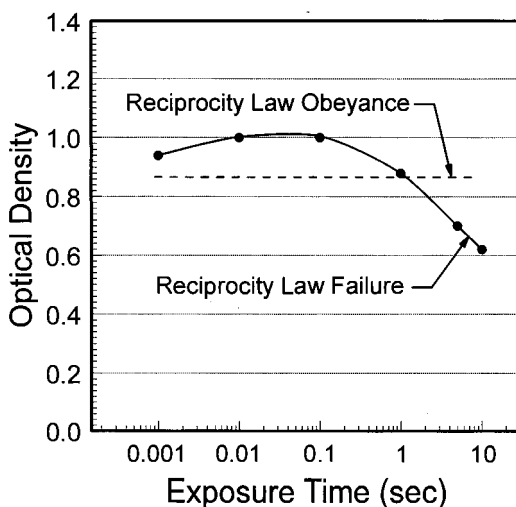


FIGURE 6-19. The concept of reciprocity law failure is illustrated. At all points along the curve, the screen-film system is exposed exactly the same but at different exposure rates in the time indicated on the x-axis. At short and long exposure times, a reduction in the efficiency or speed of the screen-film system occurs. If the reciprocity law were to hold, the curve would appear as that shown as the dashed line.

exposure times the exposure rate is low, and at short exposure times the exposure rate is high. In the extremes of exposure rate, the film becomes less efficient at using the light incident upon it, and lower ODs result. Reciprocity law failure can become a factor in mammography, where long exposure times (more than 2 seconds) can be experienced for large or dense breasts.

6.7 CONTRAST AND DOSE IN RADIOGRAPHY

The screen-film system governs the overall detector contrast. However, on a case-by-case basis, the contrast of a specific radiographic application is adjusted according to the requirements of each study. The total x-ray exposure time (which governs the potential for motion artifacts) and the radiation dose to the patient are related considerations. The size of the patient is also a factor. The technologist adjusts the subject contrast during a specific radiographic procedure by adjusting the kVp. Lower kVp settings yield higher subject contrast (see Chapter 10), especially for bone imaging. Adjusting the kVp is how the technologist adjusts the beam *quality*. Quality refers to an x-ray beam's penetration power and hence its spectral properties. However, at lower kVps, the x-ray beam is less penetrating; the mAs must be increased at low kVp and decreased at high kVp to achieve proper film density. The mAs is the product of the mA (the x-ray tube current) and the exposure time (in seconds); it is used to adjust the number of x-ray photons (beam *quantity*) produced by the x-ray system at a given kVp.

How does the technologist determine which kVp to use for a specific examination? The appropriate kVp for each specific type of examination is dogmatic, and these values have been optimized over a century of experience. The kVp for each type of radiographic examination (e.g., forearm; kidney, ureter, and bladder; cervical spine; foot; lateral hip) should be listed on a *technique chart*, and this chart should be posted in each radiographic suite. The technologist uses the recommended kVp for the radiographic examination to be taken, but usually adjusts this value depending on the size of the patient. Thicker patients generally require higher kVps. Most modern radiographic systems use automatic exposure control (i.e., phototiming), so that the time of the examination is adjusted automatically during the exposure. Phototiming circuits have a backup time that cannot be exceeded. If the kVp is set too low for a given patient thickness, the time of the exposure may be quite long, and if it exceeds the time that is set on the backup timer, the exposure will terminate. In this situation, the film will generally be underexposed and another radiograph will be required. The technologist should then increase the kVp to avoid the backup timer problem. Use of a proper technique chart should prevent excessively long exposure times. To illustrate the change in beam penetrability versus kVp, Table 6-2 lists tissue half value layers for x-ray beams produced at various kVps.

Figure 6-20 shows the typical geometry used in radiography. For a 23-cm-thick patient as an example, the entrance skin exposure (ESE, the exposure that would be measured at the surface of the skin) required to deliver an exposure of 0.6 mR to the screen-film detector system (which should produce an OD of about 1.0 for a 400-speed system) was calculated at each kVp. The table in Fig. 6-20 shows approximate ESE values versus kVp, and it is clear that lower kVps result in higher ESE values to the patient. The data shown in Fig. 6-20 were computed so that when the

TABLE 6-2. TISSUE HALF-VALUE LAYERS (HVLs) VERSUS PEAK KILOVOLTAGE (kVp) computed using water as tissue-equivalent material, assuming a 5% voltage ripple and 2 mm of added aluminum filtration

kVp	Tissue HVL (cm)
40	1.48
50	1.74
60	1.93
70	2.08
80	2.21
90	2.33
100	2.44
110	2.53
120	2.61
130	2.69
140	2.76

kVp was reduced, the mA was increased to ensure the same degree of film darkening.

Figure 6-21 illustrates the kVp dependence of the entrance dose (the dose to the first 1 cm of tissue), average tissue dose, and subject contrast for a 23-cm-thick patient. Contrast for a 1-mm thickness of bone was calculated. As with the previous figure, the data shown were computed so that the same film darkening would be delivered. Although the bone contrast is reduced with increasing kVp, the radiation dose (both entrance dose and average dose) plummets with increasing kVp. This curve illustrates the classic compromise in radiography between contrast and dose to the patient.

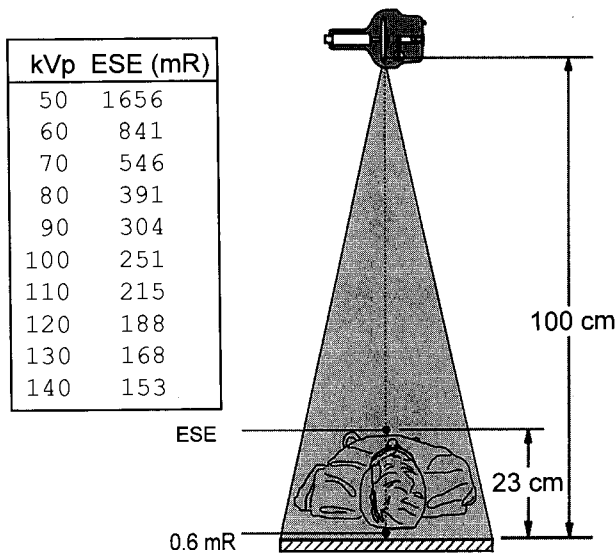


FIGURE 6-20. The geometry of a typical radiographic exposure. The entrance skin exposure (ESE) is shown as a function of peak kilovoltage (kVp) (inset). Any of the radiographic techniques shown in the table will produce the same optical density on a 400-speed screen-film system (delivering 0.6 mR to the screen-film cassette). Because lower-energy x-ray beams are less penetrating, a larger ESE is required to produce proper film darkening, which results in higher patient exposure.

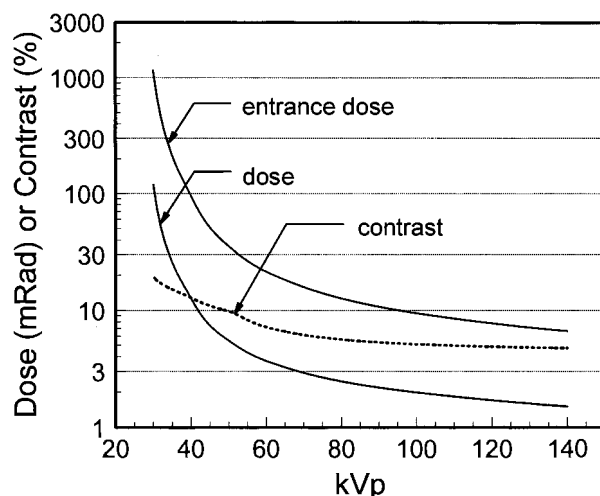


FIGURE 6-21. Contrast and dose as a function of peak kilovoltage (kVp). The contrast (in percent) of a 1-mm bone chip within a 23-cm patient is reduced as the kVp increases. Beyond 90 kVp, the reduction in contrast is slight. The entrance dose (calculated in the first 1 cm of tissue) and the average dose (averaged over the 23-cm thickness of the patient) are illustrated. Exponential attenuation of the x-ray beam results in the much lower average dose. Because the y-axis has exponential units, small vertical changes on the graph correspond to large actual changes. Lower kVps result in higher bone versus tissue contrast but also result in substantially higher x-ray doses to the patient.

6.8 SCATTERED RADIATION IN PROJECTION RADIOGRAPHY

The x-ray energies used in general radiography range from about 15 to 120 keV. In tissue, the probability of Compton scattering is about the same as for the photoelectric effect at 26 keV. Above 26 keV, the Compton scattering interaction is dominant. For materials of higher average atomic number (e.g., bone), the probability of equal contributions from photoelectric and Compton interactions occurs at about 35 keV. Therefore, for virtually all radiographic procedures except mammography, most photon interactions in soft tissue produce scattered x-ray photons.

Scattered photons are detrimental in radiographic imaging because they violate the basic geometric premise that photons travel in straight lines (see Fig. 6-1). As shown in Fig. 6-22A, the scattered photon is detected at position P on the detector, but this position corresponds to the trajectory of the dashed line, not of the scattered photons. The detection of scattered photons causes film darkening but does not add information content to the image.

How significant is the problem of scattered radiation? Fig. 6-22B illustrates the concept of the ratio of scatter (S) to primary (P) photons—the S/P ratio. The S/P ratio depends on the area of the x-ray field (field of view), the thickness of the patient, and the energies of the x-rays. Figure 6-2 illustrates the S/P ratio for various field sizes and patient thicknesses. For abdominal radiography, a 35- × 35-cm (14- × 14-inch) field of view is typically used, and even a trim adult patient is at least 20 cm thick. From Fig. 6-23, this corresponds to an S/P ratio greater than 3. For a S/P ratio of 3, for every primary photon incident on the screen-film detector, *three* scattered photons strike the detector: 75% of the photons striking the detector carry little or no useful information concerning the patient's anatomy. For larger patients, the S/P ratio in the abdomen can approach 5 or 6.

Figure 6-23 illustrates the advantage of aggressive field collimation. As the field of view is reduced, the scatter is reduced. Therefore, an easy way to reduce the amount of x-ray scatter is by collimating the x-ray field to include only the anatomy of interest and no more. For example, in radiography of the thoracic spine, collimation to just the spine itself, and not the entire abdomen, results in better radi-

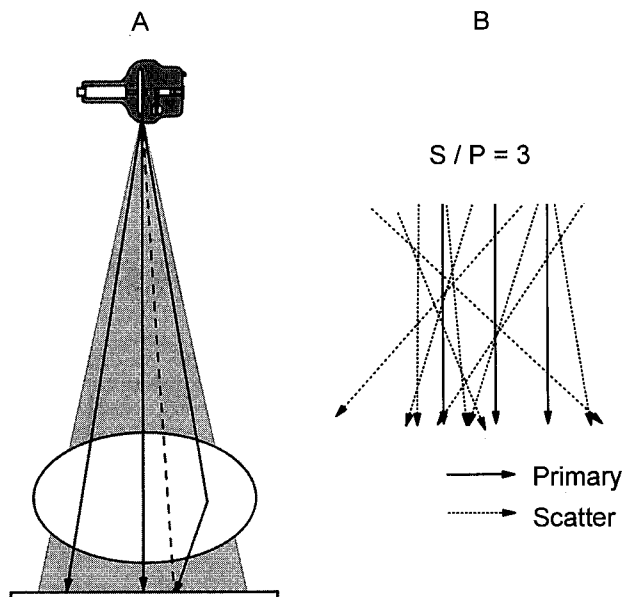


FIGURE 6-22. A: Scattered radiation violates the basic principle of projection imaging, by which photons are supposed to travel in straight lines. The scattered x-ray illustrated in the figure affects the film density at a position corresponding to the dashed line; however, the scattered photon does not carry information concerning the attenuation properties of the patient along that dashed line. **B:** The scatter-to-primary ratio (S/P ratio) refers to how many scattered x-ray photons (S) there are for every primary photon (P). Three primary photons (solid lines) are shown, along with nine scatter photons (dotted lines), illustrating the situation for $S/P = 3$.

ographic contrast of the spine. However, overcollimation can result in the occasional exclusion of clinically relevant anatomy, requiring a retake.

In screen-film radiography, scattered radiation causes a loss of contrast in the image. For two adjacent areas transmitting photon fluences of A and B, the contrast in the absence of scatter (C_0) is given by the following equation:

$$C_0 = \frac{A - B}{A}$$

For an object with contrast equal to C_0 in the absence of scatter, the contrast in the presence of scatter (C) is given by the following:

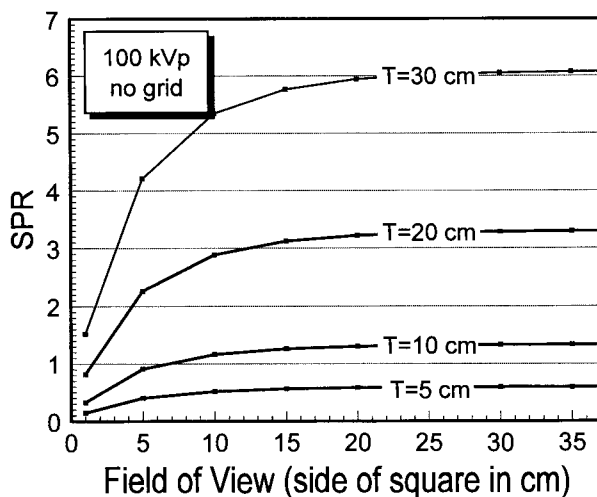


FIGURE 6-23. The scatter-to-primary ratio (S/P ratio) as a function of the side dimension of the x-ray field (i.e., the field of view) is shown for various patient thicknesses. The S/P ratios correspond to a 100-kVp exposure in the absence of an antiscatter grid. For a typical abdominal film, with a patient 20 to 30 cm thick over a 30- × 30-cm field area, an S/P between 3 and 6 will occur in the absence of a grid. Thinner patients or patient parts result in substantially reduced S/P ratios.

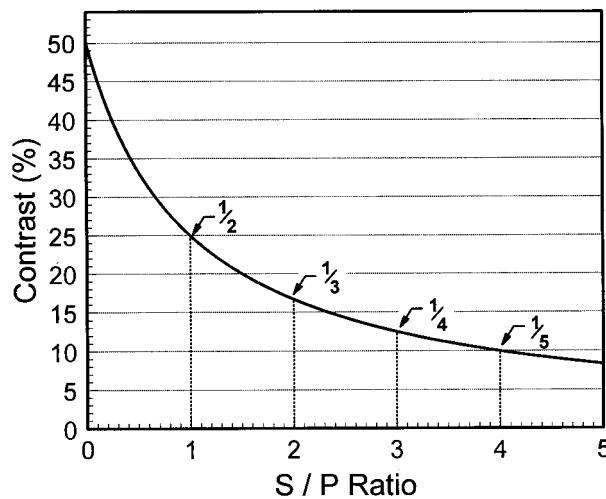


FIGURE 6-24. The contrast of an object as a function of the scatter-to-primary (S/P) ratio. An object that has 50% contrast in the absence of scatter has the contrast reduced to one-fourth of the original contrast with an S/P = 3. The shape of this curve shows the devastating impact on contrast that scatter has in screen-film radiography.

$$C = C_0 \times \frac{1}{1 + S/P}$$

This expression is plotted in Fig. 6-24, with $C_0 = 50\%$. An S/P of 1.0 reduces an object's contrast by a factor of 2, and an S/P of 2.0 reduces contrast by a factor of 3. The $(1 + S/P)^{-1}$ term is called the *contrast reduction factor*.

The Antiscatter Grid

The antiscatter grid is used to combat the effects of scatter in diagnostic radiography. The grid is placed between the patient and the screen-film cassette (Fig. 6-25). The grid uses geometry to reduce the amount of scatter reaching the detector. The source of primary radiation incident upon the detector is the x-ray focal spot, whereas the scattered radiation reaching the detector emanates from countless dif-

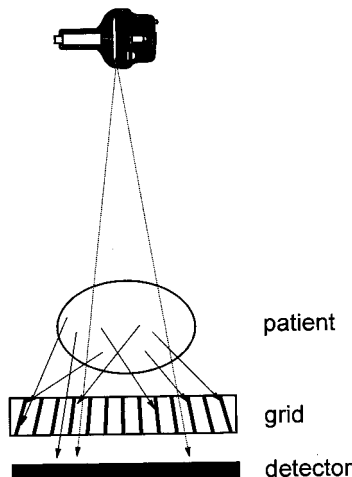


FIGURE 6-25. The geometry of an antiscatter grid used in radiography. The grid, which is composed of a series of lead grid strips that are aligned (focused) with respect to the x-ray source, is placed between the patient (the source of scatter) and the detector. More primary photons than scatter photons pass through the interspaces of the grid. Consequently, the grid acts to reduce the scatter-to-primary (S/P) ratio that is detected on the radiograph. (Note: The horizontal dimension of the grid has been exaggerated to improve detail in the illustration.)

ferent locations from within the patient. An antiscatter grid is composed of a series of small slits, aligned with the focal spot, that are separated by highly attenuating septa. The thickness of the grid in Fig. 6-25 is exaggerated to illustrate its internal structure; the typical thickness is about 3 mm, including the top and the bottom surfaces. Because the grid slits are aligned with the source of primary radiation, primary x-rays have a higher chance of passing through the slits unattenuated by the adjacent septa. The septa (grid bars) are usually made of lead, whereas the openings (interspaces) between the bars can be made of carbon fiber, aluminum, or even paper.

The typical dimensions of an antiscatter grid are shown in Fig. 6-26. There are many parameters that describe the design and performance of grids. The most clinically important parameters concerning grids are discussed here, and the remaining parameters are summarized in Table 6-3. The single most important parameter that influences the performance of a grid is the *grid ratio*. The grid ratio is simply the ratio of the height to the width of the *interspaces* (not the grid bars) in the grid. Grid ratios of 8:1, 10:1, and 12:1 are most common in general radiography, and a grid ratio of 5:1 is typical in mammography. The grid is essentially a one-dimensional collimator, and increasing the grid ratio increases the degree of collimation. Higher grid ratios provide better scatter cleanup, but they also result in greater radiation doses to the patient. A grid is quite effective at attenuating scatter that strikes the grid at large angles (where 0 degrees is the angle normal to the grid), but grids are less effective for smaller-angle scatter. The scatter cleanup as a function of scatter angle for a typical radiographic imaging situation is shown in Fig. 6-27. With no grid present, the scattered radiation incident upon the detector is broadly distributed over angles ranging from 0 to 90 degrees, with a peak toward intermediate angles. The influence of various grid ratios is shown. All grids are effective at remov-

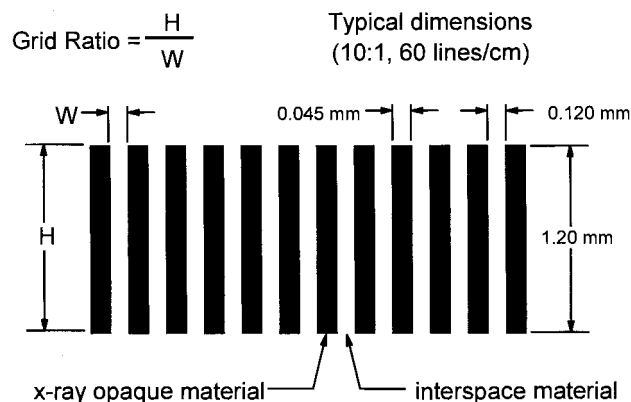


FIGURE 6-26. The details of grid construction. Parallel grid bars (typically made of lead or tantalum) and interspaces (typically made of aluminum, cardboard, or carbon fiber) correspond to an infinite-focus grid (i.e., parallel x-rays). Most grids are focused, with common focal distances of 90 to 110 cm (35 to 43 inches) for general radiography and 170 to 190 cm (67 to 75 inches) for chest radiography. The thickness of the 10:1 grid (illustrated) is 1.20 mm; however, all grids are covered on both sides with aluminum or carbon fiber for support, making the total thickness of a grid about 3 mm.

TABLE 6-3. ADDITIONAL PARAMETERS (NOT DISCUSSED IN THE TEXT) THAT DESCRIBE ANTI-SCATTER GRID PROPERTIES

Parameter	Description
Primary grid transmission	The fraction of primary x-ray photons that pass through the grid; for a perfect grid, this would be unity.
Scatter transmission	The fraction of scattered x-ray photons that pass through the grid; for a perfect grid, this would be zero.
Selectivity	The ratio of the primary transmission to the scatter transmission of the grid.
Lead content	The amount of lead in the grid, measured in units of mg/cm^2 (density $\times$ thickness).
Contrast improvement factor	The contrast achieved with the grid divided by the contrast without the grid.
Grid type	Most grids are linear grids, but crossed grids (in which the grid lines are arranged in a symmetric matrix) are available also; crossed grids are used infrequently.

ing very highly angled scattered photons (e.g., greater than 70 degrees). Grids with higher grid ratios are more effective at removing intermediate-angle scatter.

The *focal length* of the grid determines the amount of slant of the slits in the grid (see Fig. 6-25), which varies from the center to the edge. Infinite-focus grids have no slant (i.e., the interspaces are parallel). Typical grid focal lengths are 100 cm (40 inches) for general radiography or 183 cm (72 inches) for chest radiography. Grids with lower grid ratios suffer less grid cutoff if they are slightly off-focus, compared with grids that have high grid ratios. For this reason, grids used in fluoroscopy, where the source-to-detector distance changes routinely, often have a lower grid ratio.

If it is stationary, the grid will cast an x-ray shadow of small parallel lines on the image. To eliminate this, the grid can be moved in the direction perpendicular

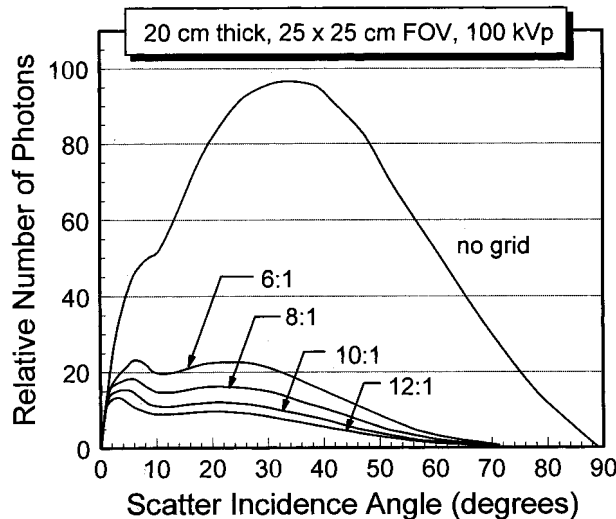


FIGURE 6-27. The distribution of scattered x-ray photons as a function of the scatter incidence angle. In the absence of a grid ("no grid"), the angle of scattered radiation, which emanates from the patient and strikes the detector system, is illustrated. In the presence of a grid, the number of scattered x-ray photons that strike the detector is markedly reduced and the incidence angle is shifted to smaller angles. Grids with higher grid ratios pass fewer scattered x-ray photons than those with lower grid ratios do.

lar to the direction of the slits. Moving the grid during the x-ray exposure blurs the grid bars so that they are not visible on the radiographic image. The device that moves the grid on a radiographic system is called the *Bucky* (after its inventor).

The *grid frequency* refers to the number of grid bars per unit length. Grids with 40 and 60 grid lines per centimeter are commonly available. Grids with 80 grid lines per centimeter are available but quite expensive. The grid frequency does not influence the scatter cleanup performance of the grid. For stationary grids, the grid frequency determines the spacing of the striped pattern that the grid produces on the film. Many institutions now make use of a stationary grid for upright chest radiography rather than bear the expense of the Bucky system, and the radiologist ignores the finely spaced stripe pattern in the image. The grid frequency is very important if the film is to be digitized or if digital receptors are used for radiography. A stationary grid imposes a periodic pattern that can cause aliasing with digital imaging systems. (Aliasing describes an image artifact that is caused by insufficient digital sampling). Aliasing often produces moiré patterns (see Chapter 10 for more information on aliasing).

The *interspace material* influences the dose efficiency of the grid. Ideally, the interspace material should be air; however, the material must support the malleable lead septa. Aluminum is the most common interspace material, although carbon fiber is becoming standard on the more expensive grids. Carbon fiber or air interspaces are an absolute requirement for the low-energy realm of mammography.

The *Bucky factor* describes one aspect of the performance of a grid under a set of clinical conditions (including kVp and patient thickness). The Bucky factor is the ratio of the entrance exposure to the patient when the grid is used to the entrance exposure without the grid while achieving the same film density. Bucky factors for various kVps and grid ratios are illustrated in Fig. 6-28, as determined by computer simulations. Bucky factors range from about 3 to 5.

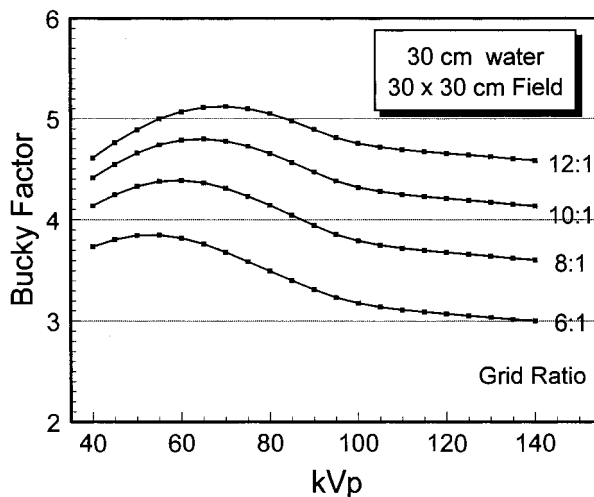


FIGURE 6-28. The Bucky factor for four different grid ratios is shown as a function of peak kilovoltage (kVp) for a 30-cm field of view and 30-cm thickness of water. The Bucky factor is the increase in the entrance exposure to the patient that is required (at the same kVp) to compensate for the presence of the grid and to achieve the same optical density. Higher grid ratios are more effective at reducing detected scatter but require a higher Bucky factor.

Artifacts Caused by Grids

Most artifacts associated with the use of a grid have to do with mispositioning. Because grids in radiographic suites are usually permanently installed, mispositioning generally occurs only during installation or servicing of the system. For portable examinations, grids are sometimes used as a *grid cap* (which allows the grid to be firmly attached to the cassette). The chance of grid mispositioning in portable radiography is much higher than in a radiographic suite. Common positioning errors associated with the use of a grid are illustrated in Fig. 6-29. The radiologist should be able to identify the presence and cause of various grid artifacts when present. For example, for an upside-down grid (see Fig. 6-29), there will be a band of maximum OD along the center of the image in the direction of the grid bars (usually vertical), and there will be significant grid cutoff (reduced OD on the film) toward both sides of the image.

Air Gaps

The use of an air gap between the patient and the screen-film detector reduces the amount of detected scatter, as illustrated in Fig. 6-30. The clinical utility of this approach to scatter cleanup is compromised by several factors, including the addi-

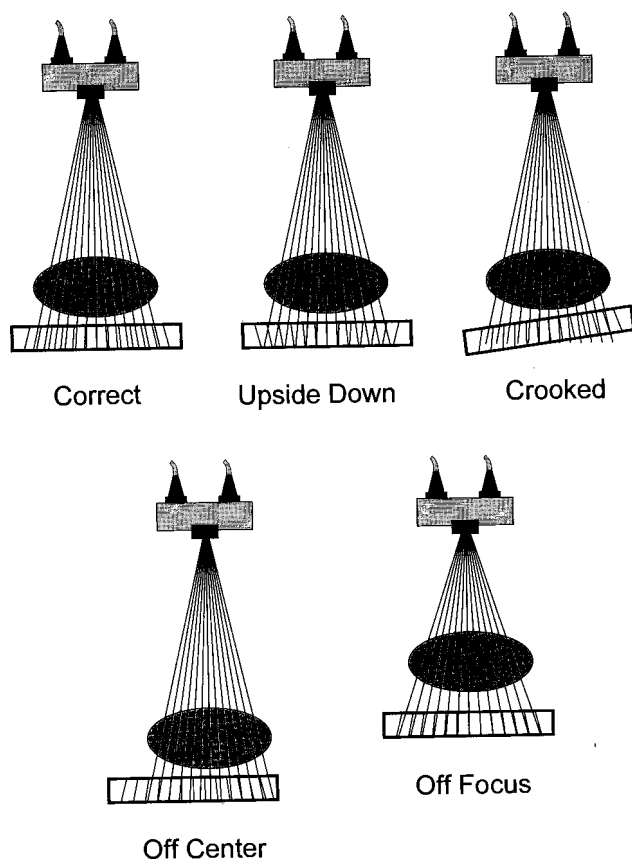


FIGURE 6-29. Grid orientation. The correct orientation of the grid with respect to the patient and x-ray source is illustrated in the upper left drawing. Four improperly positioned grids are illustrated in the subsequent diagrams.

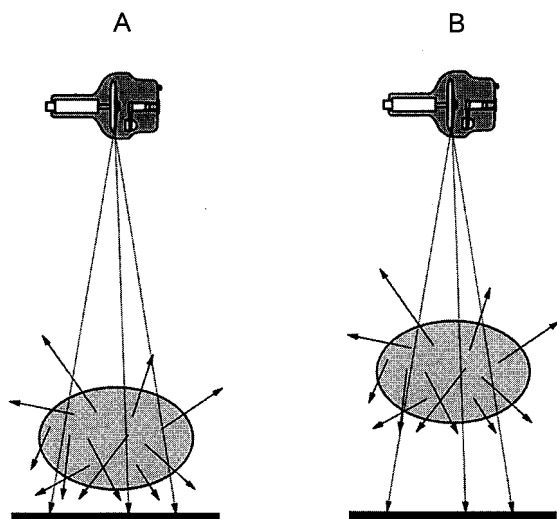


FIGURE 6-30. Air gap geometry can reduce the amount of scatter radiation reaching the detector system. When the distance between the patient and the detector is increased (**B**), fewer x-ray photons reach the detector system compared with the positioning in (**A**), owing to simple geometric considerations.

tional object magnification and the reduced field of view that is imaged by a detector of fixed dimensions. The additional object magnification often causes a loss of spatial resolution, unless a very small focal spot is used. In general, air gaps have not enjoyed much use in general radiography for scatter cleanup, outside of chest radiography. In chest radiography, the larger SID allows a small air gap to be used (e.g., 20 cm) without imposing too much magnification. Grids are still routinely used in all of radiography other than extremity studies, some pediatric studies, and magnification mammography views.

SUGGESTED READING

Haus AG. The AAPM/RSNA physics tutorial for residents: Measures of screen-film performance. *Radiographics* 1996;16:1165–1181.

FILM PROCESSING

Film is a unique technology, born of the 19th century, which allows an optical scene to be recorded. Film itself was a novelty when x-rays were discovered in 1895. In the early days of radiography, film emulsion was coated onto glass plates and exposed to x-rays, and prints from the developed plates were produced for interpretation. No intensifying screen was used. With the start of World War I in 1914, the supply of glass plates from Belgium was halted, while the demand for medical radiography for injured soldiers increased. A cellulose nitrate film base with a single-sided emulsion was developed and used with a single intensifying screen to produce radiographs. By 1918, a dual emulsion film was developed and dual-screen cassettes were used, which allowed a reduction of the exposure time. Cellulose nitrate was highly flammable, and it was succeeded by less flammable cellulose triacetate. In the 1960s, polyester was introduced and became the standard film base in the radiography industry.

Film processing was performed by hand, using a system of chemical vats, until the first automatic film processor (which was more than 3 m long and weighed about 640 kg) was introduced in 1956. This system could produce a dry radiographic film in 6 minutes. Although the use of film is gradually giving way to digital technologies in the 21st century, film remains a high-resolution recording medium that serves as the optical detector, the display device, and the medical archive.

7.1 FILM EXPOSURE

Film Emulsion

The film emulsion consists of silver halide crystals suspended in a high-quality, clear gelatin. A thin adhesive layer attaches the film emulsion to the flexible polyester base. The silver halide is approximately 95% AgBr and 5% AgI. The positive silver atoms (Ag^+) are ionically bonded to the bromine (Br^-) and iodine (I^-) atoms. Silver halide forms a cubic crystalline lattice of AgBr and AgI. The silver halide crystals used for film emulsion have *defects* in their lattice structure introduced by silver sulfide, AgS. Defects in crystalline structure give rise to unique chemical properties in many substances. For example, a small amount of terbium is added to $\text{Gd}_2\text{O}_2\text{S}$ ($\text{Gd}_2\text{O}_2\text{S}:\text{Tb}$) to initiate defects in that crystal, and this “activator” is largely responsible for the excellent fluorescent properties of $\text{Gd}_2\text{O}_2\text{S}:\text{Tb}$ intensifying screens. Europium is added to BaFBr ($\text{BaFBr}:\text{Eu}$) to introduce lattice defects in that storage

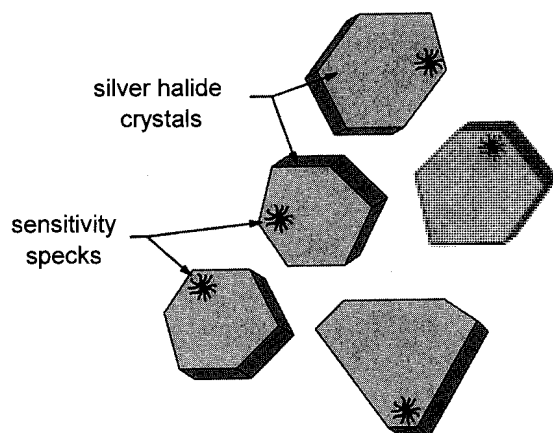


FIGURE 7-1. A conceptual illustration of tabular grain silver halide crystals is illustrated. The *sensitivity speck* is a place on the crystal where silver reduction takes place; for exposed crystals, the sensitivity speck becomes the location of the *latent image center*.

phosphor. The defect in the silver halide crystal gives rise to its optical properties in the region of a *sensitivity speck* (Fig. 7-1).

In the silver halide crystal, the Ag atoms and the halide atoms share electrons in an ionic bond, and a negative charge (the excess electrons contributed by Br^- and I^-) builds up on the surface of the silver halide crystal. A positive charge (Ag^+) accumulates inside the crystal. The sensitivity speck, induced by defects in the crystal, is essentially a protrusion of positive charge that reaches the surface of the crystal (Fig. 7-2).

The Latent Image

The hundreds of visible light photons emitted by each x-ray interaction in the intensifying screen are scattered through the screen matrix, and some fraction reach the interface between the screen and the film emulsion. Reflection and refraction takes place at this interface, and a number of the light photons cross the interface

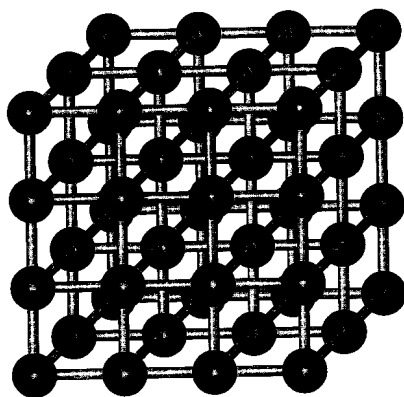


FIGURE 7-2. The cubic lattice of a silver halide crystal is illustrated.

and eventually reach the film emulsion. Light photons are little packets of energy, and loosely bound electrons in the silver halide can absorb light energy and drift as free electrons in the emulsion. The negatively charged free electrons can come in contact with the positive sensitivity speck, where ionic silver (Ag^+) is *reduced* to metallic silver by the free electron:



Recall that the loss of an electron is called oxidation; the gain of an electron is called reduction. Thus, light exposure causes a small number of the silver ions in the film emulsion to become reduced and be converted to metallic silver, Ag. It is known that a two-atom Ag complex has increased thermal stability over single-atom Ag. The formation of the two-atom Ag center is called the *nucleation phase* of latent image formation. The nucleation center acts as a site for attracting additional free electrons and mobile silver ions, forming more Ag atoms during the *growth phase* of latent image formation. Experimental evidence indicates that a minimum of three to five reduced silver atoms are needed to produce a *latent image center*. Silver halide crystals that have latent image centers will subsequently be developed (further reduced to metallic silver by chemical action) during film processing, whereas silver halide crystals that lack latent image centers will not be developed. A film that has been exposed but not yet developed is said to possess a *latent image*. The latent image is not visible but is encoded in the film emulsion as the small clusters of Ag atoms in some of the silver halide crystals.

Development

After exposure, the film emulsion is subjected to a sequence of chemical baths, which leads to *development*. Silver halide crystals that have not been exposed in the emulsion, and therefore do not possess a latent image center (i.e., a localized deposit of metallic silver atoms), are inert to the effects of the chemical developer and are unchanged. They are ultimately washed out of the emulsion. For silver halide crystals that have been exposed to light photons, the metallic silver atoms at the latent image center act as a chemical *catalyst* for the developer. During the development process, the latent image center catalyzes the reaction, which reduces the remaining silver ions in that silver halide crystal into a grain of metallic silver (Fig. 7-3). Each blackened grain contributes very slightly to the optical density (OD) of the film at that location. The contribution of millions of microscopic blackened photographic grains is what gives rise to the gray scale in the image. Darker areas of the film have a higher concentration of grains (per square millimeter), and lighter areas have fewer grains. The grains are about 1 to 5 μm in size and are easily visible under a light microscope. Once chemically developed, the latent image (which is invisible) becomes a *manifest image*.

Development is a chemical amplification process, and an impressive one at that. Amplification factors of 10^9 are experienced as the silver atoms in the latent image center catalyze the chemical development of additional silver atoms in the silver halide.

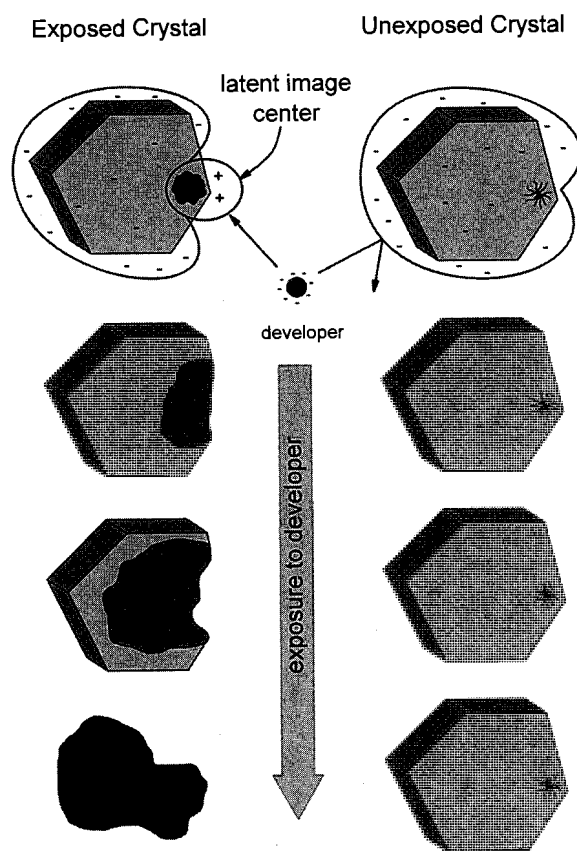


FIGURE 7-3. The fate of an exposed and an unexposed silver halide crystal is illustrated through the development stage. Exposed silver halide crystals possess a latent image center, which is a collection of at least 3 to 5 Ag atoms, and typically have 10 or more Ag atoms. The Ag atoms act to catalyze the further reduction of the silver in the silver halide crystal in the presence of the reducing agent in the chemical developer solution. After development, exposed silver halide crystals become black silver grains, and unexposed silver halide crystals are washed out of the emulsion and discarded.

7.2 THE FILM PROCESSOR

A modern automatic film processor is illustrated in Fig. 7-4. The film is fed into the input chute of the processor (inside the darkroom) and shuttled through the processor by a system of rollers. The film is guided by the rollers through the developer tank, into the fixer tank, and then into the wash tank. After washing, the film continues its path through the dryer, where temperature-regulated air is blown across the film surface and dries it.

One of the functional requirements of the automatic film processor is to deliver consistent performance, film after film, day after day. Chemical reactions in general, and certainly those involved in film development, are highly dependent on both temperature and chemical concentrations. Therefore, the processor controls both the temperature and the concentration of the processing chemicals. A thermostat regulates the temperature of each of the solution tanks to ensure consistent development. The temperature of the developer tank is most crucial, and typically this temperature is 35°C (95°F). Pumps circulate the liquid in each tank to ensure adequate thermal and chemical mixing.

As sheets of film are run through the processor, the reactions that occur between the chemicals in each tank and the film emulsion act to deplete the concentrations of some of the chemicals. The automatic film processor replenishes

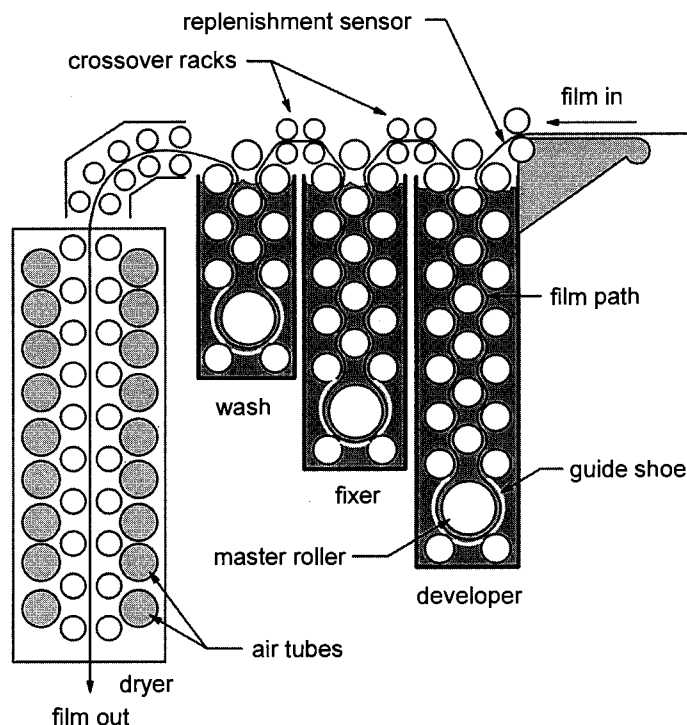


FIGURE 7-4. The functional components of a film processor are illustrated. Film follows the circuitous path through each of the three tanks and then through the dryer. The rollers are driven at the same rotation rate by a single motor and usually a chain mechanism is used to tie in all the rollers. At the bottom of each tank, the film direction changes direction 180°, and a larger roller with a guide shoe is used for this. Crossover racks transfer the film from tank to tank.

developer and fixer chemicals by pumping in fresh developer and fresh fixer from storage tanks. When a sheet of film is fed into the input chute of the processor, it passes by and trips a small switch that activates the replenishment pumps. For each 35 cm of film length, about 65 mL of developer and 100 mL of fixer are pumped into the appropriate tanks by the processor.

In addition to temperature and concentration, the extent of a chemical reaction also depends on how long it is allowed to occur. The film passes through the roller assembly at a constant speed, and all of the rollers are driven (usually by a chain and sprocket assembly) by a single-speed regulated motor. The time that the emulsion spends in each of the chemical baths is regulated by the length of the film path through each tank, which is governed by the depth of the tank and by the film speed.

Development

The developer solution includes water, developing agents, an activator, a restrainer, a preservative, and a hardener (Table 7-1). The developing agents are reducing agents, and hydroquinone and phenidone are often used. These chemicals act as a source of

TABLE 7-1. DEVELOPER CHEMICAL SUMMARY

Agent	Chemical	Purpose
Developer	Hydroquinone and Phenidone	Reducing agent, causes $\text{Ag}^+ + e^- \rightarrow \text{Ag}$ reaction
Activator	Sodium hydroxide	Drives pH up for development reactions
Restrainer	KBr and KI	Limits reaction to exposed grains only
Preservative	Sodium sulfite	Prevents oxidation of developer
Hardener	Glutaraldehyde	Controls emulsion swelling

electrons to initiate the $\text{Ag}^+ + e^- \rightarrow \text{Ag}$ reaction. This reaction is the “development” of the film. The phenidone participates in chemical reactions in areas of the film that contain a low concentration of latent image centers (which will be the lighter, lower OD regions of the manifest image), and the hydroquinone participates more in areas of the emulsion that are more heavily exposed (darker areas of the image).

The *activator* is sodium hydroxide, a strong base that serves to increase the pH of the developer solution. The activator also causes the gelatin in the emulsion to swell, which provides better solute access to the silver halide crystals. Potassium halides (KBr and KI) act as a *restrainer*, which helps to limit the development reaction to only those silver halide crystals that have been exposed and have latent image centers. The restrainer puts enough halide ions in the development solution to adjust the reaction kinetics appropriately. A *preservative* such as sodium sulfite helps to prevent oxidation of the developer chemicals by airborne oxygen. Finally, a *hardener* is used to control the swelling of the emulsion. Too much swelling can cause excessive softening of the gelatin, resulting in water retention. Glutaraldehyde is used as the hardener.

Periodically, the chemicals from the processor are dumped and new chemicals are added. Typically a fresh batch of developer processes too “hot”; that is, the films are initially be too dark. To reduce this problem, *starter* is added to the developer tank, which acts to *season* the chemicals.

Fixing

As the film is transported out of the developer solution, a pair of pinch rollers wrings most of the developer solution out of the wet emulsion. A small amount of developer remains trapped in the emulsion and would continue to cause development unless neutralized. The first role of the fixer solution is that of a stop bath—to arrest the chemical activity of the residual developer. Because the developer thrives in a basic solution, the fixer contains acetic acid to decrease the pH and stop the development. In the fixer, the acetic acid is called the *activator* (Table 7-2). Ammonium thiosulfate is used in the fixer as a *clearing agent*. In the film processing industry, the clearing agent is also referred to as *hypo*. The clearing agent removes the undeveloped silver halide crystals from the emulsion. *Hypo retention* refers to the undesirable retention of excessive amounts of ammonium thiosulfate in the emulsion, which leads to poor long-term stability of the film. A *hardener* such as potassium chloride (KCl) is included in the fixer solution to induce chemical hardening of the emulsion. The hardening process shrinks and hardens the emulsion, which helps the emulsion to be dried. Like the developer, the fixer also contains sodium sulfite, which acts as a *preservative*.

TABLE 7-2. FIXER CHEMICAL SUMMARY

Agent	Chemical	Purpose
Activator	Acetic acid	Stops developer reaction by driving pH down
Clearing agent (hypo)	Ammonium thiosulfate	Removes undeveloped silver halide from emulsion
Hardener	Potassium chloride	Shrinks and hardens emulsion
Preservative	Sodium sulfite	Prevents oxidation of fixer

Washing

The water bath simply washes residual chemicals out of the film emulsion. Water is continuously cycled through the processor; room temperature fresh water comes in from the plumbing to the wash tank, and water is drawn away from the tank and fed to a drain.

Drying

The last step that the film experiences as it is transported through the processor is drying. A thermostatically regulated coil heats air blown from a powerful fan, and this warm air blows across both surfaces of the film. A series of air ducts distributes the warm air evenly over both surfaces of the film to ensure thorough drying.

7.3 PROCESSOR ARTIFACTS

Hurter and Driffield Curve Effects

Some forms of film processing artifacts can be demonstrated on the Hurter and Driffield (H & D) curve. If the concentration of the developer is too high, overdevelopment may occur (Fig. 7-5). With overdevelopment, the higher ODs cannot be

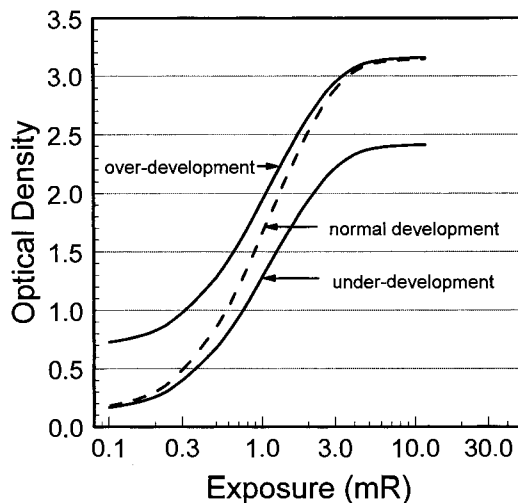


FIGURE 7-5. The influence of overdevelopment and underdevelopment is illustrated on the Hurter and Driffield (H & D) curve.

increased too much because the number of silver grains present in the emulsion limits these OD values. Overdevelopment therefore primarily affects the low OD part of the H & D curve, where silver halide crystals that were not exposed to light become developed.

Underdevelopment occurs if the developer concentration is too low or if the temperature of the developer is too low. Underdevelopment manifests as a reduction in OD at the high exposure end of the H & D curve. In addition to changes in density that are seen with overdevelopment or underdevelopment (Fig. 7.5), the slope of the H & D curve is reduced and the films will appear flat and lack contrast.

Other Artifacts

Water spots can occur on the film if the replenishment rates are incorrect, if the squeegee mechanism that the film passes through after the wash tank is defective, or if the dryer is malfunctioning. Water spots are best seen with reflected light. This and other artifacts are shown conceptually in Fig. 7-6. A *slap line* is a plus-density line perpendicular to the direction of film travel that occurs near the trailing edge of the film. This is caused by the abrupt release of the back edge of the film as it passes through the developer-to-fixer crossover assembly. *Pick-off* artifacts are small, clear areas of the film where emulsion has flecked off from the film base; they can be caused by rough rollers, nonuniform film transport, or a mismatch between the film emulsion and chemicals. *Wet pressure marks* occur when the pinch rollers apply too much or inconsistent pressure to the film in the developer or in the developer-to-fixer crossover racks. Several types of *shoe marks* can occur. Shoe marks result when the film rubs against the guide shoes during transport. The artifact manifests as a series of evenly spaced lines parallel the direction of film transport. Plus-density shoe marks are caused by the

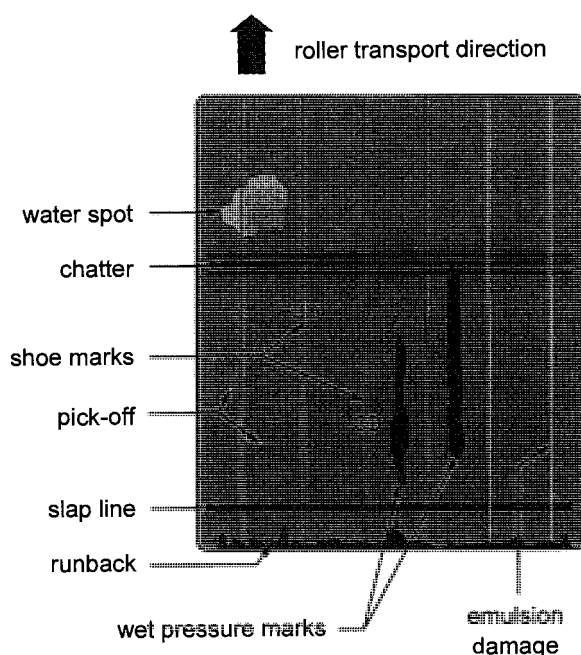


FIGURE 7-6. Typical artifacts associated with film processors are illustrated.

guide shoes in the developer tank. Minus-density shoe marks often occur in the fixer-to-washer crossover. If emulsion surface damage is present, the problem can be anywhere along the transport path. *Runback* artifacts look like fluid drips and occur at the trailing edge of a film. In this situation, the film transport rollers from the developer to the fixer tank do not remove developer from the film surface (no squeegee action), and as the film descends into the fixer the excess developer “runs back” at the trailing edge of the film. Developer on the developer-to-fixer crossover assembly becomes oxidized and can cause this artifact. *Chatter* is a periodic set of lines perpendicular to the film transport direction that is caused by binding of the roller assembly in the developer tank or in the developer-to-fixer crossover assembly.

7.4 OTHER CONSIDERATIONS

Rapid Processing

The standard film process in radiology requires 90 seconds to process the film, but in some areas of a hospital (typically emergency room radiographic suites), 1½ minutes is too long to wait for the film. The roller transport in the film processor can be sped up by changing gear ratios: by a simple 2:1 reduction gear, a 45-second processor transport time can be achieved. When the transport time is decreased, the chemical activity of the development process must be increased. This can be accomplished by increasing the concentration of the developer, by increasing the temperature of the developer, or both. Typically, the processor temperature is increased to about 38°C for a 45-second process.

Extended Processing

Extended processing, also called *push processing*, is the act of slowing down the film transport, typically increasing the transport time to 120 seconds. Extended processing received interest in mammography, where it results in a slightly faster system (Fig. 7-7) and a commensurate reduction in the radiation dose to the patient

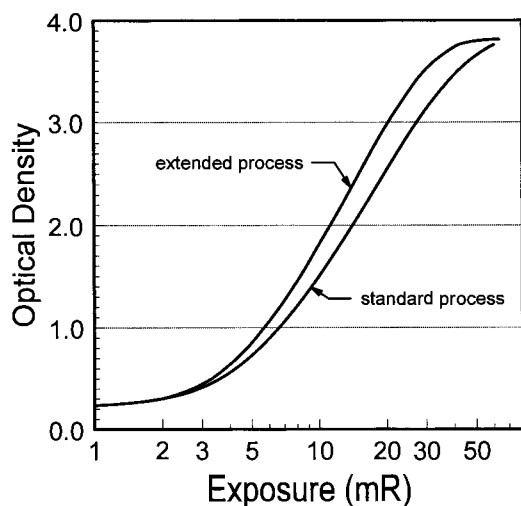


FIGURE 7-7. Extended or “push” processing refers to extending the processing time; it is used to increase the speed of the screen-film system. The processing time can be extended by changing gear ratios in the roller drive system of the film processor.

can be achieved. Film manufacturers have responded by designing screen-film mammography systems with higher sensitivity, and therefore the popularity of extended processing has waned.

Daylight Processing

Technologist productivity is an important factor in a busy department, and in larger institutions it is inefficient for the technologist to load and unload the screen-film cassette by hand. To increase throughput, daylight processing systems are often used. Daylight processors are automatic mechanical systems connected to a film processor. Exposed cassettes are fed into the slot of the system by the technologist in a lighted room, the cassette is retracted into a dark chamber, the exposed film is removed from the cassette and fed into the film processor, and a fresh sheet of film is loaded into the cassette. The cassette then pops out of the slot, back into the hands of the technologist.

7.5 LASER CAMERAS

Laser cameras, also called laser imagers, are used to expose sheets of film, using a set of digital images as the source. Laser cameras are typically used to produce films from computed tomography (CT) or magnetic resonance imaging (MRI) systems. Once exposed in the laser camera, the film is processed in a standard wet chemistry film processor. Some laser cameras have film processors attached, and these combination systems expose the film and then pass it to the processor automatically. Other laser cameras can expose films and deliver them into a light-tight film magazine, which can be taken to a film processor where the films are processed.

Film designed for use in laser cameras typically uses cubic grain technology, instead of the tabular grain film used in screen-film cassettes (Fig. 6-13 in Chapter 6). Film is exposed in a laser camera with a laser beam, which scans the surface of the film. A computer controls the position and intensity of the laser beam as it is scanned in raster-fashion across the surface of the film (Fig. 7-8). The intensity of the laser source itself is not modulated to adjust the film density; rather, an optical modulator sits in the laser beam path and is used to rapidly adjust the laser intensity as the beam scans the film. The light in the laser beam exposes the film in the same manner as light from an intensifying screen, and the process of latent image center formation is identical to that described previously.

7.6 DRY PROCESSING

Dry processors, like laser cameras, are used for producing images from digital modalities, such as ultrasound, digital radiography, CT, and MRI. The principal benefits of dry processing systems are that they do not require plumbing, they do not require chemical vats, which produce fumes, and they generate less chemical waste. However, the cost per sheet of film is higher compared to wet processing systems. For low-throughput areas where a need exists to produce film images from digital modalities but the total number of images is low, dry processors provide a cost-effective alternative to the traditional wet chemistry processor.

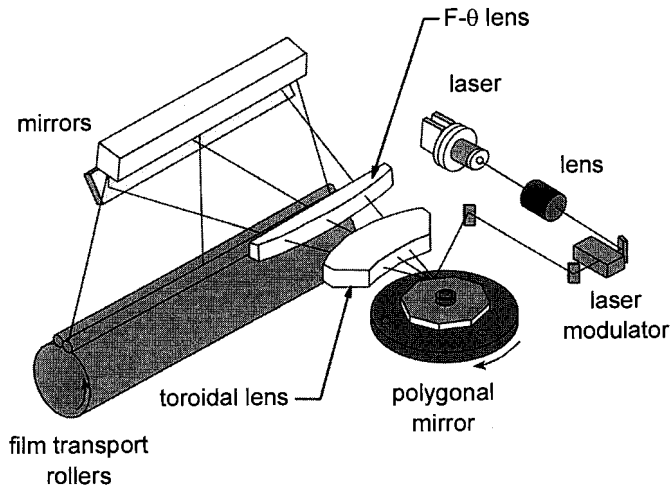


FIGURE 7-8. The optical and mechanical components of a laser camera are shown. The laser emits a constant beam, which is focused and then modulated in intensity by the laser modulator. This controls the density on the film by controlling the amount of laser energy that strikes a particular location. The laser beam strikes a rotating polygonal mirror, and rotation of the mirror at a constant rate causes the laser beam to be scanned laterally across the film, for example in the x-direction. The F-θ and toroidal mirrors convert angular motion of the incident laser beam to linear motion across the surface of the film. The film is transported through the system by the rollers, and this motion is in the y-direction. Using these components, film is optically exposed to the digital image by a combination of selecting the spot location and modulating the laser beam intensity.

There are several dry processing technologies that are available from different manufacturers. One dry system makes use of carbon instead of silver grains to produce density on the film. This technique is not photographic, but *adherographic*. The raw film consists of an imaging layer and a laser-sensitive adhesive layer, sandwiched between two polyester sheets. The imaging layer is a matrix of carbon particles and polymer. As the laser scans the film, laser energy is focused at the laser-sensitive adhesive layer. Thermal sensitization by the laser causes adhesion of the carbon layer to the polyester film base. After laser exposure, the two outer polyester layers are separated, producing a positive and a negative image. The positive image is then coated with a protective layer and becomes the readable film, and the negative layer is discarded. All of these steps are performed in the dry laser camera, and the finished image pops out of the system.

Because the adhesion process is binary, gray scale is produced using a dithering pattern. For example, each area on the film corresponding to a pixel in the image is segmented into a number of very small (approximately $5\ \mu\text{m}$) *pels*, and 16×16 pels per pixel can be used to produce 256 different shades of gray.

Another dry processor makes use of silver behenate and silver halide, and the scanning laser energy provides thermal energy, which drives the development of the silver halide crystals. The gray scale is a “true” gray scale, so a dither pattern is not necessary. However, the undeveloped silver halide crystals remain in the film layer, and subsequent heating (e.g., storage in a hot warehouse) can cause image stability problems.

7.7 PROCESSOR QUALITY ASSURANCE

Modern x-ray equipment is computer controlled, with feedback circuits and fault detection capability. Laser cameras are also computer controlled and highly precise. Because of their chemical nature and the huge degree of amplification that film processing performs, x-ray processors are typically the least precise instruments in the generation of diagnostic images. Therefore, for quality control (QC) in the radiology department, the processors are a “first line of defense.” Quality control of film processors is a standard of practice that is scrutinized by the Joint Commission on the Accreditation of Health Organizations (JCAHO) and regulatory bodies.

The tools required for processor QC are fairly basic. A *sensitometer* (Fig. 7-9) is a device that exposes film (in the darkroom) to a standardized range of light exposures. To expose the film to various exposure levels, the sensitometer shines light through an optical step tablet, which is a sheet of film that has different density steps on it. The sensitometer is used to expose film, which is used in the department, and these films are then processed and the densities are measured on the film. To measure the OD on a film, a *densitometer* (Fig. 7-10) is used. To use the densitometer, a film is placed in the aperture of the device, the readout button is depressed, and the OD is displayed on the densitometer. A densitometer can be used in standard room lighting conditions.

The processor quality assurance program consists of a technician who exposes a series of filmstrips in the darkroom each morning, and then takes these filmstrips to each processor in the department and runs them. The processed filmstrips are then collected, and the OD at three or four different steps is measured and recorded

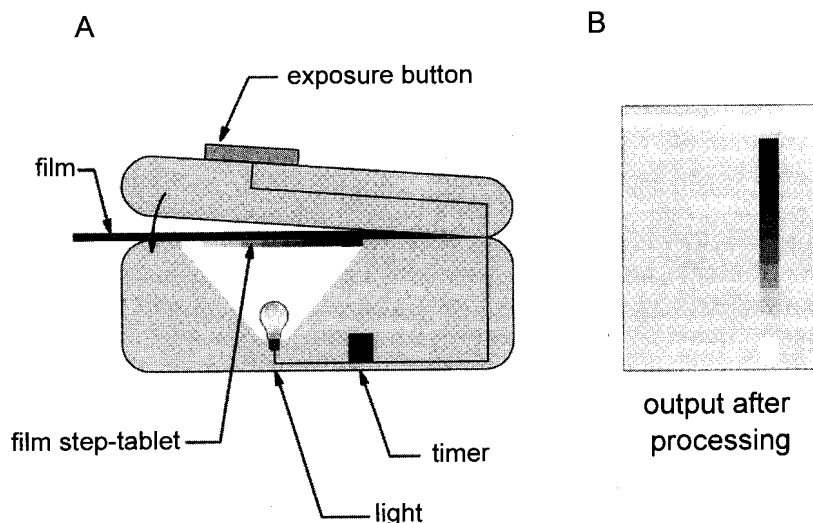


FIGURE 7-9. A: A schematic diagram of a film sensitometer is shown. The sensitometer is used to expose, in a darkroom, a sheet of raw film to a stepped series of different light intensities. **B:** After the film is processed, a step-edge of different optical densities (ODs) results from the sensitometer exposure. Sensitometers can be used to acquire the data to plot a *relative* Hurter and Driffield (H & D) curve.

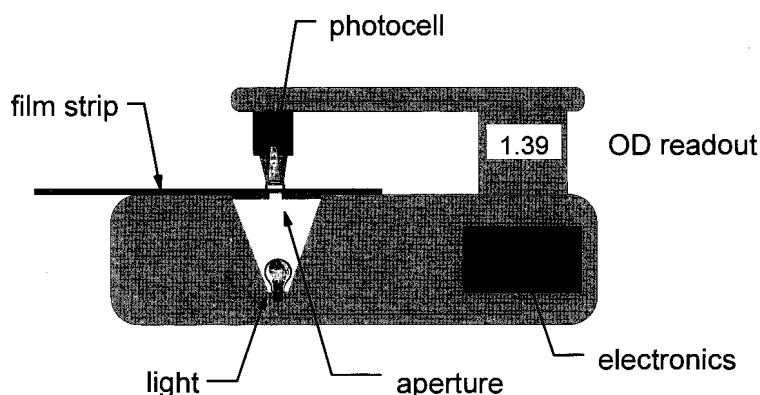


FIGURE 7-10. The functional components of a densitometer are illustrated. A densitometer is used to measure the optical density (OD) at a particular location on a piece of film. For a more thorough discussion of OD, see Chapter 6.

on a monthly log for each processor (Fig. 7-11). The *base + fog* reading is simply the OD measured anywhere on the processed film (away from any of the exposed steps). The *speed* index is the OD measured at one of the mid-gray steps on the density strip. Usually the step corresponding to an OD of about 1.0 + base + fog is selected. The *contrast* index is the difference in OD between two of the steps near the mid-gray region of the stepped exposures. The purpose of processor quality assurance is to recognize problems with processor performance *before* those problems cause a diagnostic exposure to be compromised. This is why processor QC should be done first thing in the morning, before the clinical work schedule begins.

Measuring the performance of each processor on a daily basis will do nothing for the clinical image quality unless there is an action plan when things go wrong. For processor QC, the three parameters measured (base + fog, speed, and contrast) each should have an *action limit*. Action limits are illustrated in Fig. 7.11 as horizontal dashed lines. If the measured index exceeds (in either direction) the acceptable range of performance, then the processor should be taken out of service and the processor maintenance personnel should be called. Action limits should be set realistically, so that real problems get fixed and minor fluctuations do not stimulate unnecessary responses. The action limits are set by monitoring a given processor over some period of time. Experience and statistical techniques are used to set appropriate action limits. Action limits for older and lower-throughput processors, for example, are often set more leniently (wider). The evaluation of processor response using sensitometry can be done manually, or commercially available systems can be used to semiautomatically measure, record, and plot the sensitometry data. Daily processor QC is considered optimal, although some institutions reduce the processor monitoring frequency to two or three times per week.

Different types of film have different responses to small changes in chemistry: some films are largely impervious to small processor changes, and others have a more pronounced response. It is best to use the film that is most sensitive to changes in processing as the film for processor QC. Radiographic screens, depending on the

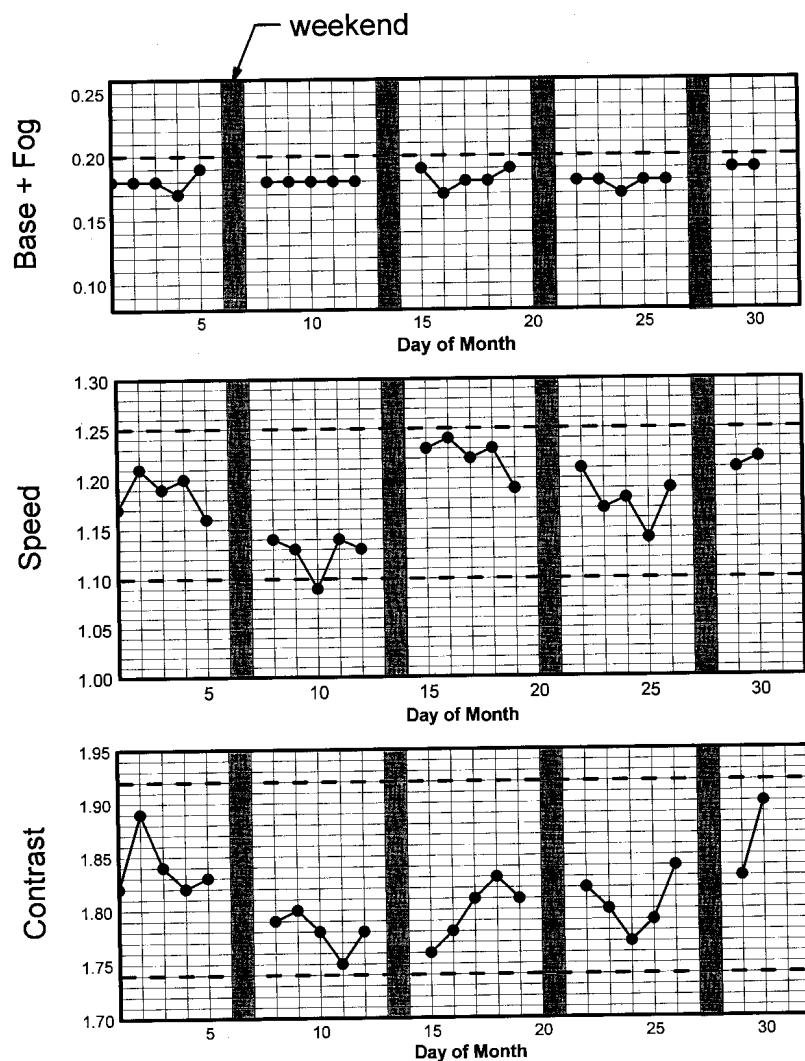


FIGURE 7-11. Three charts typically used for monitoring film processor performance are shown. Daily measurements (excluding weekends) are plotted. The base + fog level is the background optical density (OD) of the film. The speed index is simply the OD of a mid-gray step, and changes in this value indicate a lateral shift in the Hurter and Driffield (H & D) curve. The contrast index is the numerical difference between two steps, and it is related to the gradient or slope of the H & D curve. Trends in each parameter can be used to spot processor problems before they affect diagnostic images. The horizontal dashed lines are the *action limits*; when any parameter extends beyond the action limit (as shown for the speed index on day 10), appropriate action should be taken to fix the problem.

manufacturer and the phosphor employed, emit light in different spectral regions (usually in the blues or greens), and most laser cameras expose the film using red light. Some sensitometers allow color filter selection to most closely approximate the spectral conditions experienced clinically.

Each film company manufactures its film based on anticipated demand to keep the film fresh, and each box of film is marked with an *emulsion number* corresponding to the manufacturing session. Small differences in film speed are expected between different boxes of film. Processor QC is a program designed to measure the reproducibility (i.e., precision) of the *processor*. Therefore, the same box of film should be used for processor QC so that changes caused by differences in emulsion numbers do not affect the results. When the box of QC is almost empty, a new box should be opened and films from both boxes should be used in the processor QC procedure for about 3 to 5 days of overlap. This should be sufficient to track differences from one box of film to the next.

SUGGESTED READING

- Bogucki TM. Laser cameras. In: Gould RG and Boone JM, eds. *Syllabus. A categorical course in physics: technology update and quality improvement of diagnostic x-ray imaging equipment*. Oak Brook, IL: RSNA, 1996:195–202.
- Bushong SC. *Radiological sciences for technologists: physics, biology, and protection*, 5th ed. St. Louis: Mosby, 1993.
- Dainty JC, Shaw R. *Image science*. London: Academic Press, 1974.
- Haus AG, ed. *Film processing in medical imaging*. Madison, WI: Medical Physics Publishing, 1993.
- James TH. *The theory of the photographic process*, 4th ed. Rochester, NY: Eastman Kodak, 1977.

MAMMOGRAPHY

Mammography is a radiographic examination that is specially designed for detecting breast pathology. Interest in breast imaging has been fostered by the realization that approximately one in eight women will develop breast cancer over a lifetime. Technologic advances over the last several decades have greatly improved the diagnostic sensitivity of mammography. Early x-ray mammography was performed with nonscreen, direct exposure film, required high radiation doses, and produced images of low contrast and poor diagnostic quality. In fact, it is possible that early mammography screening exams, in the 1950s and 1960s, did not provide any useful benefits for the early detection of breast cancer. Mammography using the xeroradiographic process was very popular in the 1970s and early 1980s, spurred by good spatial resolution and edge-enhanced images; however, the relatively poor contrast sensitivity coupled with a higher radiation dose in comparison to the technologic advances of screen-film mammographic imaging led to the demise of the Xerox process in the late 1980s. Continuing refinements in technology have vastly improved mammography over the last 15 years, as illustrated by breast images from the mid-1980s and late 1990s (Fig. 8-1). The American College of Radiology (ACR) mammography accreditation program changed the practice of mammography in the mid-1980s, with recommendations for minimum standards of practice and quality control, which spurred technologic advances and ensured quality of service. In 1992, the federal Mammography Quality Standards Act (MQSA) largely adopted the voluntary program and incorporated many of the recommendations into regulations. The goal of the MQSA is to ensure that all women have access to quality mammography for the detection of breast cancer in its earliest, most treatable stages, with optimal patient care and follow-up. The MQSA regulations continue to evolve and adapt to changing technologies. Emerging full-field digital mammography devices promise rapid image acquisition and display with better image quality, anatomy-specific image processing, and computer-aided detection tools to assist the radiologist in identifying suspicious features in the images.

Breast cancer screening programs depend on x-ray mammography because it is a low-cost, low-radiation-dose procedure that has the sensitivity to detect early-stage breast cancer. Mammographic features characteristic of breast cancer are masses, particularly ones with irregular or "spiculated" margins; clusters of microcalcifications; and architectural distortions of breast structures. In screening mammography, two x-ray images of each breast, in the mediolateral oblique

and craniocaudal views, are routinely acquired. Early detection of cancer gives the patient the best chance to have the disease successfully treated. As a result, the American Medical Association, American Cancer Society, and American College of Radiology recommend yearly mammograms for women throughout their lives, beginning at age 40. While screening mammography attempts to identify breast cancer in the asymptomatic population, *diagnostic* mammography is performed to further evaluate abnormalities such as a palpable mass in a breast or suspicious findings identified by screening mammography. Diagnostic mammography includes additional x-ray projections, magnification views, and spot compression views. X-ray stereotactic biopsy or studies using other imaging modalities may also be performed.

Of specific interest for augmenting diagnostic mammography are ultrasound, magnetic resonance imaging (MRI), and, in some cases, nuclear medicine imaging. Ultrasound (Chapter 16) has a major role in differentiating cysts (typically benign) from solid masses (often cancerous), which have similar appearances on the x-ray mammogram, and in providing biopsy needle guidance for extracting breast tissue specimens. MRI (Chapters 14 and 15) has tremendous tissue contrast sensitivity, is useful for evaluating silicone implants, and can accurately assess the stage of breast cancer involvement. Although the role of complementary imaging modalities in breast evaluation and diagnosis is small compared to x-ray screen-film mammography, significant advances may lead to increased use in the future.

Same breast imaged 10 years apart

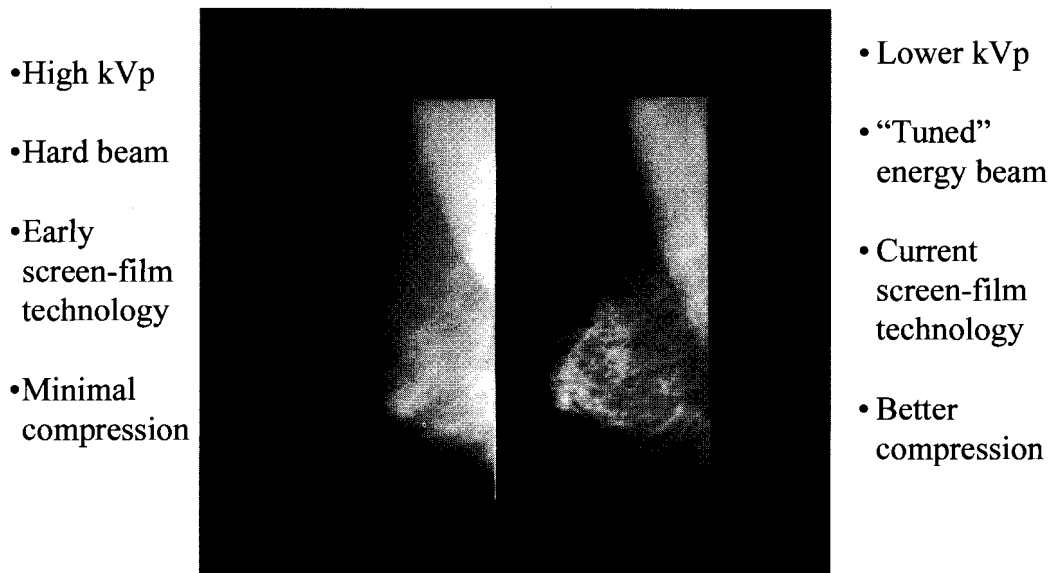


FIGURE 8-1. Improvements in mammography image quality over the last 10 years are exemplified by a mid-1980s image and an image taken approximately 10 years later. Contrast and detail are significantly improved.

Other modalities that have been used for breast imaging include thermography, transillumination, and diaphanography, which employ the infrared or visible light portion of the electromagnetic spectrum. To date, these devices have not shown adequate sensitivity and specificity for the detection of breast cancers.

The small x-ray attenuation differences between normal and cancerous tissues in the breast require the use of x-ray equipment specifically designed to optimize breast cancer detection. In Fig. 8-2A the attenuation differences between normal tissue and cancerous tissue is highest at very low x-ray energies (10 to 15 keV) and is poor at higher energies (>35 keV). Subject contrast between the normal and malignant tissues in the breast is depicted in Fig. 8-2B. Low x-ray energies provide the best differential attenuation between the tissues; however, the high absorption results in a high tissue dose and long exposure time. Detection of minute calcifications in breast tissues is also important because of the high correlation of calcification patterns with disease. Detecting microcalcifications while minimizing dose and enhancing low-contrast detection imposes extreme requirements on mammographic equipment and detectors. Because of the risks of ionizing radiation, techniques that minimize dose and optimize image quality are essential, and have led to the refinement of dedicated x-ray equipment, specialized x-ray tubes, compression devices, antiscatter grids, phototimers, and detector systems for mammography (Fig. 8-3). Strict quality control procedures and cooperation among the technologist, radiologist, and physicist ensure that acceptable breast images are achieved at the lowest possible radiation dose. With these issues in mind, the technical and physical considerations of mammographic x-ray imaging are discussed in this chapter.

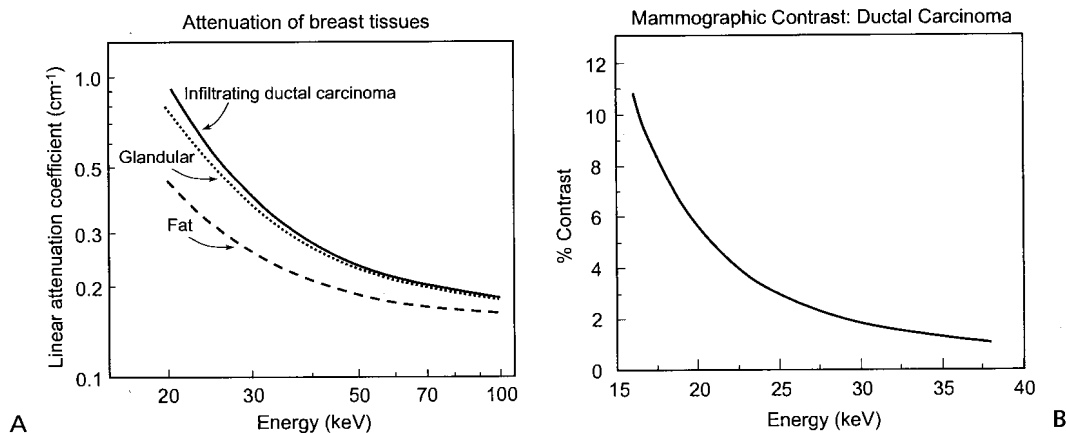


FIGURE 8-2. A: Attenuation of breast tissues as a function of energy. Comparison of three tissues shows a very small difference between the glandular and cancerous tissues that decreases with x-ray energy. **B:** Percent contrast of the ductal carcinoma declines rapidly with energy, necessitating the use of a low-energy, nearly monoenergetic spectrum. (Adapted from Yaffe MJ. Digital mammography. In: Haus AG, Yaffe MJ, eds. *Syllabus: a categorical course in physics: technical aspects of breast imaging*. Oak Brook, IL: RSNA, 1994:275-286.)

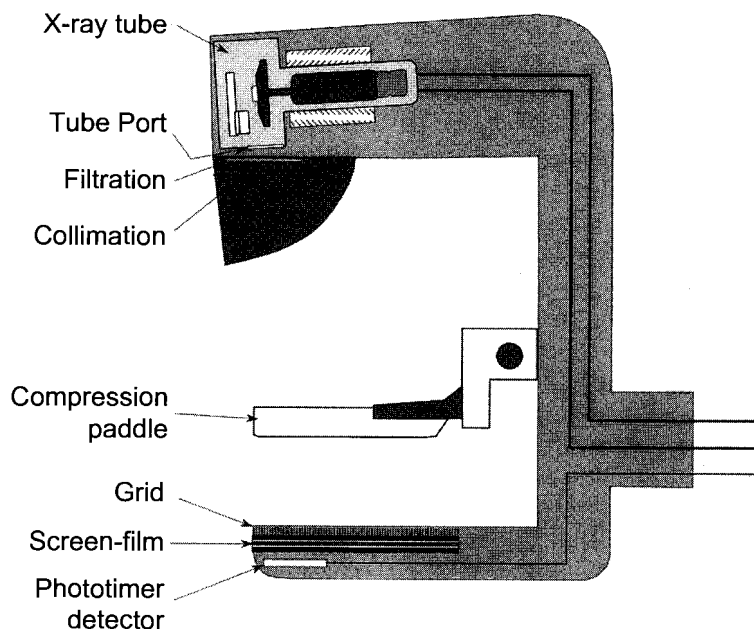


FIGURE 8-3. A dedicated mammography system has many unique attributes. Components of a typical system (excluding the generator and user console) are shown.

8.1 X-RAY TUBE DESIGN

Cathode and Filament Circuit

The mammographic x-ray tube is typically configured with dual filaments in a focusing cup that produce 0.3 and 0.1 mm nominal focal spot sizes. A small focal spot minimizes geometric blurring and maintains spatial resolution necessary for microcalcification detection. An important difference in mammographic tube operation compared to conventional radiographic operation is the low operating voltage, below 35 kilovolt peak (kVp). The *space charge effect* (Chapter 5) causes a non-linear relationship between the filament current and the tube current. Feedback circuits adjust the filament current as a function of kV to deliver the desired tube current, which is 100 mA (± 25 mA) for the large (0.3 mm) focal spot and 25 mA (± 10 mA) for the small (0.1 mm) focal spot.

Anode

Mammographic x-ray tubes use a rotating anode design. Molybdenum is the most common anode material, although rhodium and tungsten targets are also used. Characteristic x-ray production is the major reason for choosing molybdenum and rhodium. For molybdenum, characteristic radiation occurs at 17.5 and 19.6 keV, and for rhodium, 20.2 and 22.7 keV.

As with conventional tubes, the anode disk is mounted on a molybdenum stem attached to a bearing mounted rotor and external stator, and rotation is accom-

plished by magnetic induction. A source to image distance (SID) of 65 cm requires the effective anode angle to be at least 20 degrees to avoid field cutoff for the 24×30 cm field area (a shorter SID requires even a larger angle). X-ray tube anode angles vary from about 16 to 0 degrees to -9 degrees. The latter two anode angles are employed on a tube with the focal spots on the edge of the anode disk as shown in Fig. 8-4A. The effective anode angle in a mammography x-ray tube is defined as the angle of the anode relative to the horizontal tube mount. Thus for a tube with a 0-degree and -9 -degree anode angle, a tube tilt of 24 degrees results in an *effective* anode angle of 24 degrees for the large (0.3 mm) focal spot, and 15 degrees for the small (0.1 mm focal spot). In other mammography x-ray tubes, the typical anode angle is about 16 degrees, so the tube is tilted by about 6 degrees to achieve an effective anode angle of ~ 22 degrees (Fig. 8-4B). A small anode angle allows increased milliamperage without overheating the tube, because the actual focal spot size is much larger than the projected (effective) focal spot size.

The lower x-ray intensity on the anode side of the field (heel effect) at short SID is very noticeable in the mammography image. Positioning the cathode over the chest wall of the patient and the anode over the nipple of the breast (Fig. 8-5) achieves better uniformity of the *transmitted* x-rays through the breast. Orientation of the tube in this way also decreases the equipment bulk near the patient's head for easier positioning.

Mammography tubes often have *grounded anodes*, whereby the anode structure is maintained at ground (0) voltage and the cathode is set to the highest negative voltage. With the anode at the same voltage as the metal insert in this design, *off-focus radiation* is reduced because the metal housing attracts many of the rebounding electrons that would otherwise be accelerated back to the anode.

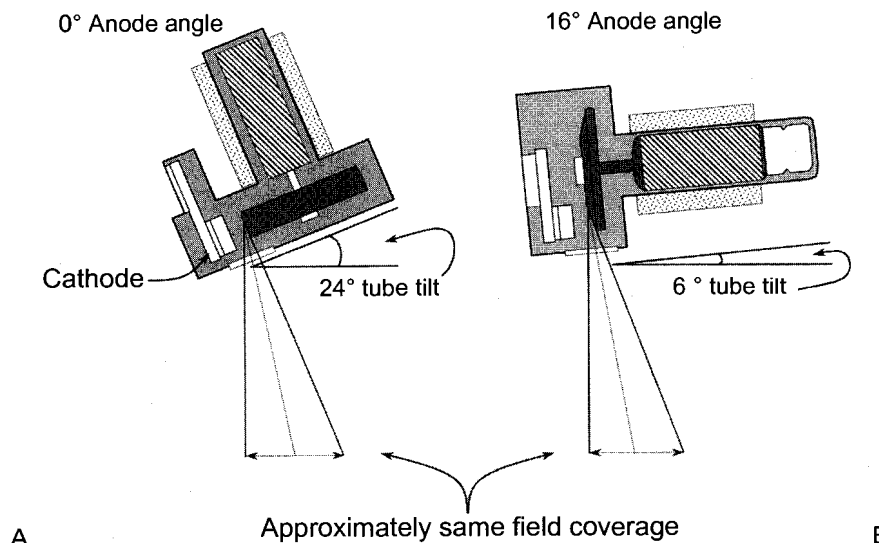


FIGURE 8-4. Design of the dedicated mammography system includes unique geometry and x-ray tube engineering considerations, including the beam central axis position at the chest wall, and a tube “tilt” angle necessary to provide adequate coverage for the 60- to 70-cm typical source to image distance. **A:** An anode angle of 0 degrees requires a tube tilt of 24 degrees. **B:** An anode angle of 16 degrees requires a tube tilt of 6 degrees.

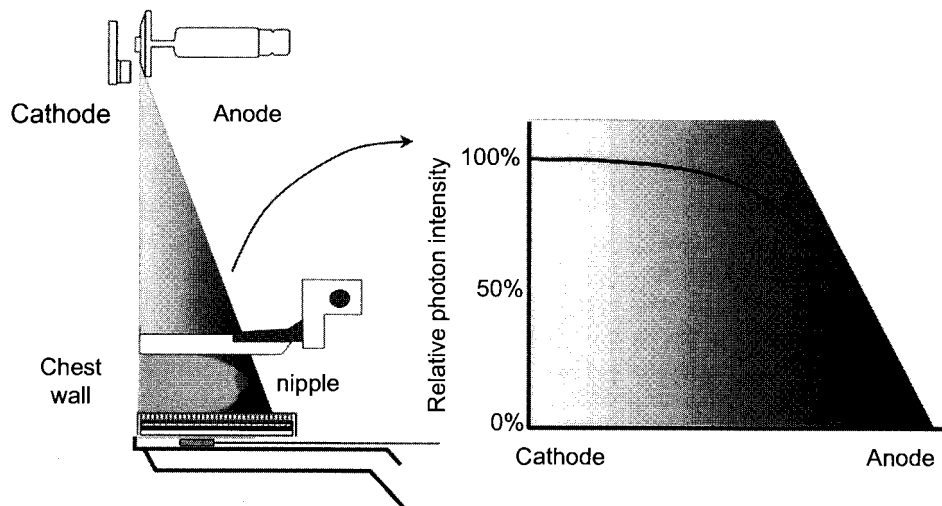


FIGURE 8-5. Orientation of the cathode-anode axis is along the chest wall to nipple direction. The heel effect, a result of anode self-filtration, describes a significant falloff of intensity toward the anode side of the x-ray beam. For this reason, the chest wall side of the breast (the thicker part of the breast) is positioned over the cathode to help equalize x-ray transmission.

Focal Spot Considerations

The high spatial resolution requirements of mammography demand the use of very small focal spots for contact and magnification imaging to reduce geometric blurring. Focal spot sizes range from 0.3 to 0.4 mm for nonmagnification (contact) imaging, and from 0.1 to 0.15 mm for magnification imaging. For a SID of 65 cm or less, a 0.3-mm nominal size is necessary for contact imaging. With longer SIDs of 66 cm or greater, a 0.4-mm focal spot is usually acceptable, because a longer SID reduces magnification and therefore reduces the influence of the focal spot size on resolution.

The focal spot and central axis are positioned over the chest wall at the receptor edge. A *reference axis*, which typically bisects the field, is used to specify the projected focal spot dimensions (Fig. 8-6). In the figure, θ represents the sum of the target angle and tube tilt, and ϕ represents the reference angle, measured between the anode and the reference axis. The focal spot length (denoted as a) at the reference axis, a_{ref} , is less than the focal spot length at the chest wall, $a_{\text{chest wall}}$, where the relationship is

$$a_{\text{ref}} = a_{\text{chest wall}} \left(1 - \frac{\tan(\theta - \phi)}{\tan \theta} \right) \quad [8-1]$$

The focal spot length varies in size with position in the image plane along the cathode-anode direction. The effective focal spot length is larger and causes more geometric blurring at the chest wall than on the anode side of the field (toward the nipple). This is particularly noticeable in magnification imaging where sharper detail is evident on the nipple side of the breast image. A *slit camera* or *pinhole camera* focal spot image measures the width and length. When the focal spot length is measured at the central axis (chest wall), application of Equation 8-1 can be used to calculate the nominal focal spot length at the reference axis. Table 8-1 indicates the nominal focal spot size and tolerance limits for the width and length.

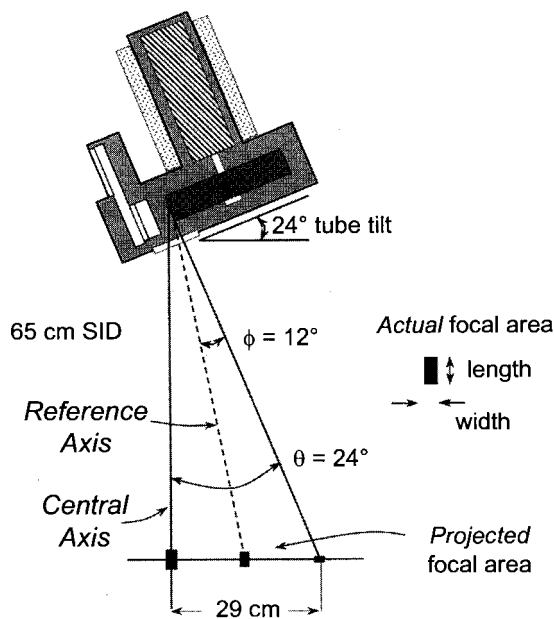


FIGURE 8-6. The projected focal spot size varies along the cathode–anode axis. The nominal focal spot size is specified at a reference axis at an angle ϕ from a line perpendicular to the cathode–anode axis. The actual size is the electron distribution area on the anode (width $\times$ length). The projected length is foreshortened according to the line focus principle. The effective focal spot length, measured on the central axis, is related to the effective focal spot length at the reference axis by Equation 8-1.

Other means to measure the effective focal spot resolution and thereby estimate the focal spot size are a star pattern (0.5-degree spokes) or a high-resolution bar pattern. The bar pattern (up to 20 line pairs/mm resolution) is required by the MQSA to measure the *effective* spatial resolution of the imaging system; the bar pattern should be placed 4.5 cm above the breast support surface. The bar pattern measurements must be made with the bars parallel and perpendicular to the cathode–anode axis to assess the effective resolution in both directions. These effective resolution measurements take into account *all of the components* in the imaging chain that influence spatial resolution (e.g., screen-film, tube motion, and focal spot size).

Tube Port, Tube Filtration, and Beam Quality

Computer modeling studies show that the optimal x-ray energy to achieve high subject contrast at the lowest radiation dose would be a monoenergetic beam of 15 to 25 keV, depending on the breast composition and thickness. Polychromatic x-rays produced in the mammography x-ray tube do not meet this need because the low-energy x-rays in the bremsstrahlung spectrum deliver significant breast dose

TABLE 8-1. NOMINAL FOCAL SPOT SIZE AND MEASURED TOLERANCE LIMITS

Nominal Focal Spot Size (mm)	Width (mm)	Length (mm)
0.10	0.15	0.15
0.15	0.23	0.23
0.20	0.30	0.30
0.30	0.45	0.65
0.40	0.60	0.85
0.60	0.90	1.30

with little contribution to the image, and the higher energy x-rays diminish subject contrast.

The optimal x-ray energy is achieved by the use of specific x-ray tube target materials to generate characteristic x-rays of the desired energy (e.g., 17 to 23 keV), and x-ray attenuation filters to remove the undesirable low- and high-energy x-rays in the bremsstrahlung spectrum. Molybdenum (Mo), ruthenium (Ru), rhodium (Rh), palladium (Pd), silver (Ag), and cadmium (Cd) generate characteristic x-rays in the desirable energy range for mammography. Of these elements, *molybdenum* and *rhodium* are used for mammography x-ray tube targets, producing major characteristic x-ray peaks at 17.5 and 19.6 keV (Mo) and 20.2 and 22.7 keV (Rh) (see Chapter 5, Table 5-2). The process of bremsstrahlung and characteristic radiation production for a molybdenum target is depicted in Fig. 8-7.

The tube port and added tube filters also play a role in shaping the mammography spectrum. Inherent tube filtration must be extremely low to allow transmission of all x-ray energies, which is accomplished with ~1-mm-thick beryllium (Be, $Z = 4$) as the tube port. Beryllium provides both low attenuation (chiefly due to the low Z) and good structural integrity.

Added tube filters of the *same* element as the target reduce the low- and high-energy x-rays in the x-ray spectrum and allow transmission of the characteristic x-ray energies. Common filters used in mammography include a 0.03-mm-thick molybdenum filter with a molybdenum target (Mo/Mo), and a 0.025-mm-thick rhodium filter with a rhodium target (Rh/Rh). A molybdenum target with a rhodium filter (Mo target/Rh filter) is also used. (Note, however, that a Mo filter cannot be used with an Rh target.) The added Mo or Rh filter absorbs the x-ray

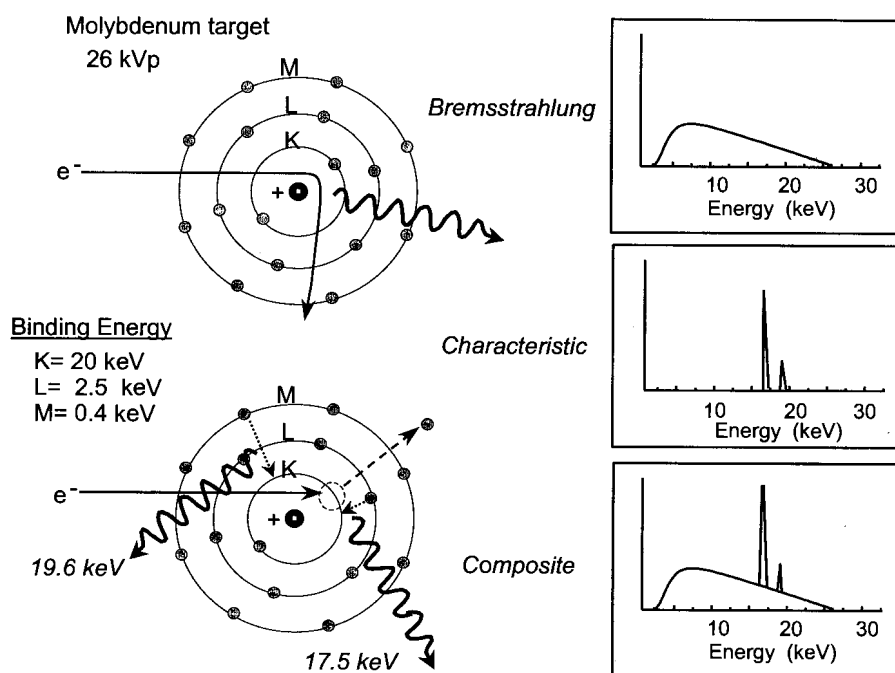


FIGURE 8-7. The output of a mammography x-ray system is composed of bremsstrahlung and characteristic radiation. The characteristic radiation energies of molybdenum (17.5 and 19.6 keV) are nearly optimal for detection of low-contrast lesions in breasts of 3- to 6-cm thickness.

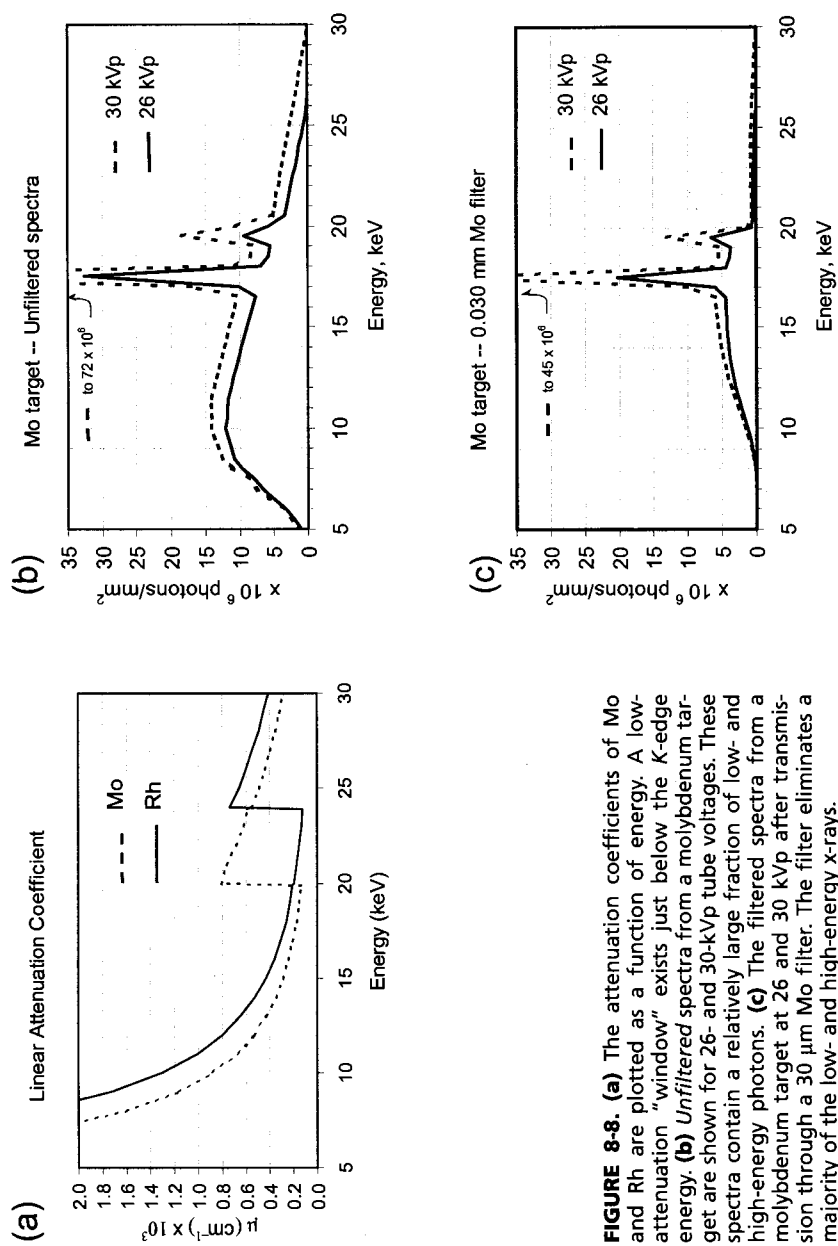


FIGURE 8-8. (a) The attenuation coefficients of Mo and Rh are plotted as a function of energy. A low-attenuation "window" exists just below the K-edge energy. **(b)** Unfiltered spectra from a molybdenum target are shown for 26- and 30-kVp tube voltages. These spectra contain a relatively large fraction of low- and high-energy photons. **(c)** The filtered spectra from a molybdenum target at 26 and 30 kVp after transmission through a 30 μ m Mo filter. The filter eliminates a majority of the low- and high-energy x-rays.

energies not useful in forming the image. Attenuation by the filter lessens with increasing x-ray energy just below the K -shell absorption edge, providing a transmission window for characteristic x-rays and bremsstrahlung photons. An abrupt increase in the attenuation coefficient occurs just above the K -edge energy, which significantly reduces the highest energy bremsstrahlung photons in the spectrum (Fig. 8-8A). Figure 8-8 shows the x-ray spectra, before and after filtration by 0.030 mm of Mo, from a molybdenum target tube operated at 26 and 30 kVp.

A molybdenum target and 0.025-mm rhodium filter are now in common use for imaging thicker and denser breasts. This combination produces a slightly higher effective energy than the Mo/Mo target/filter, allowing transmission of x-ray photons between 20 and 23 keV (Fig. 8-9). Some mammography machines have rhodium targets in addition to their molybdenum targets. A rhodium target, used with a rhodium filter, generates higher characteristic x-ray energies (20.2 and 22.7 keV), as shown by 30-kVp unfiltered and filtered spectra (Fig. 8-10). The lower melting point of Rh reduces the focal spot heat loading by approximately 20% compared to the Mo target (e.g., 80 mA_{max} for Rh versus 100 mA_{max} for Mo), so shorter exposure times are not usually possible with Rh/Rh, despite higher beam penetrability. A molybdenum filter with a rhodium target (Rh/Mo) should *never* be used, since the high attenuation beyond the K -edge of the Mo filter occurs at the characteristic x-ray energies of Rh (Fig. 8-10, *heavy dashed curve*). Comparison of output spectra using a 0.025-mm Rh filter with a Mo and Rh target (Fig. 8-11) shows the shift in the characteristic energies.

An x-ray tube voltage 5 to 10 kV above the K -edge (e.g., 25 to 30 kVp) enhances the production of characteristic radiation relative to bremsstrahlung. Characteristic radiation comprises approximately 19% to 29% of the *filtered* output spectrum when operated at 25 and 30 kVp, respectively, for Mo targets and Mo filters (see, for example, Fig. 8-8C.)

Tungsten (W) targets, used with Be tube ports and Mo and Rh filters, are available on some mammography units. The increased bremsstrahlung production efficiency of W relative to Mo or Rh allows higher power loading, and K -edge filters can shape the output spectrum for breast imaging (without any *useful* characteristic radiation). This permits higher milliamperage and shorter exposure times. However, the unfiltered tungsten spectrum (Fig. 8-12A) contains undesirable L x-rays from the W target in the 8- to 12-keV range. Attenuation of these characteristic x-rays to tolerable levels requires a Rh filter thickness of 0.05 mm compared to the 0.025 mm thickness used with Mo and Rh targets (Fig. 8-12B). With introduction of digital detectors, W anodes with added filters having K -absorption edges in the mammography energy range will become prevalent, including Mo, Rh, Pd, Ag, and Cd, among others.

Half Value Layer

The half value layer (HVL) of the mammography x-ray beam is on the order of 0.3 to 0.45 mm Al for the kVp range and target/filter combinations used in mammography (Fig. 8-13). In general, the HVL increases with higher kVp and higher atomic number targets and filters. HVLs are also dependent on the thickness of the compression paddle. For a Mo filter/Mo target at 30 kVp, after the beam passes through a Lexan compression paddle of 1.5 mm, the HVL is about 0.35 mm Al. This corresponds to a HVL in breast tissue of approximately 1 to 2 cm, although the exact

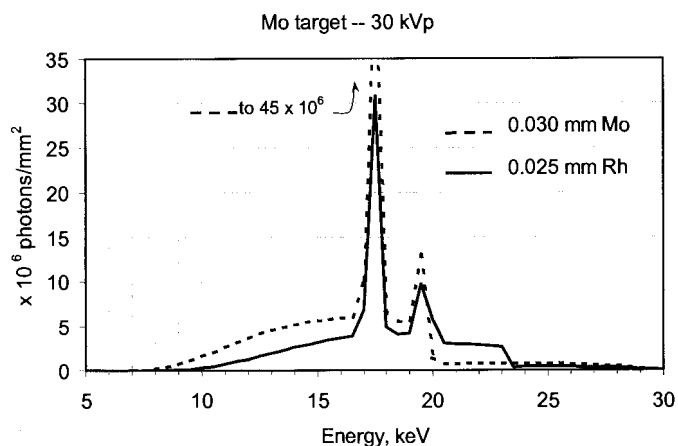


FIGURE 8-9. Output spectra from a Mo target for a 30-kVp tube voltage with a 0.030-mm Mo filter and 0.025-mm Rh filter show the relative bremsstrahlung photon transmission "windows" just below their respective K-edges.

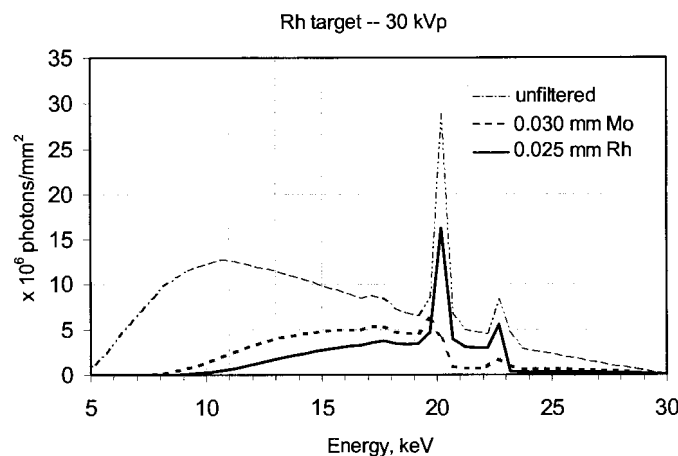


FIGURE 8-10. The rhodium target provides characteristic x-ray energies 2 to 3 keV higher than the corresponding Mo target. The unfiltered spectrum (*light dashed line*) is modified by 0.030-mm Mo (*heavy dashed line*) and 0.025-mm Rh (*solid line*) filters. A Mo filter with an Rh target inappropriately attenuates Rh characteristic x-rays.

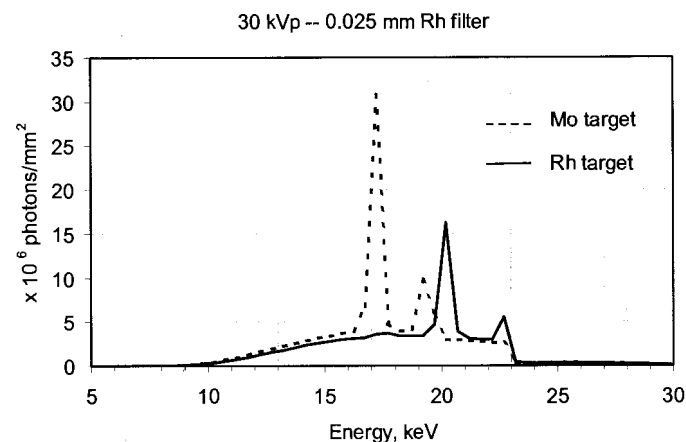


FIGURE 8-11. Spectra from beams filtered by 0.025 mm Rh are depicted for a Mo and an Rh target. The major difference is the higher energies of the characteristic x-rays generated by the Rh target, compared to the Mo target.

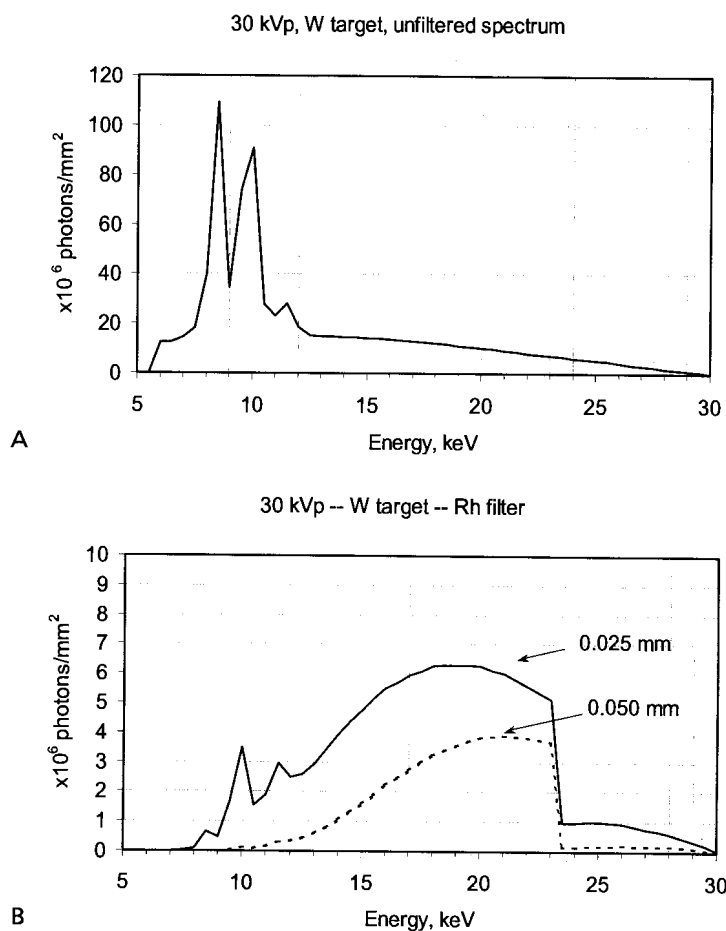


FIGURE 8-12. A: The unfiltered spectrum from a tungsten target shows a large fraction of characteristic L x-rays between 8 and 10 keV. **B:** Filtration of the beam requires approximately 0.050 mm Rh to attenuate these L x-rays to acceptable levels.

value strongly depends on the composition of the breast tissue (e.g., glandular, fibrous, or fatty tissue).

Minimum HVL limits are prescribed by the MQSA to ensure that the lowest energies in the unfiltered spectrum are removed. These limits are listed in *Table 8-2*. When the HVL exceeds certain values, for example, 0.40 mm Al for a Mo/Mo

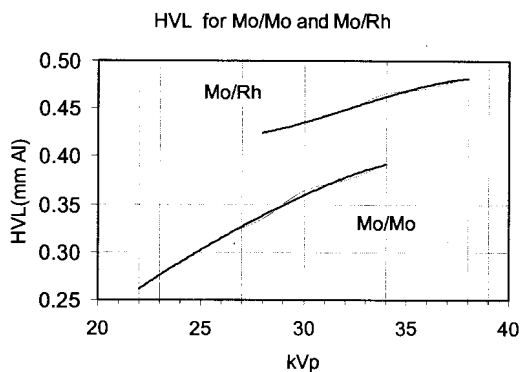


FIGURE 8-13. The half value layer versus kVp for a Mo target tube with 0.030-mm Mo filter (*bottom curve*) and 0.025-mm Rh filter (*top curve*) are plotted as a function of kVp. These values include the attenuation of the compression paddle.

TABLE 8-2. REQUIREMENTS FOR MINIMUM HALF VALUE LAYER (HVL) MQSA: 21 CFR (FDA REGULATIONS) PART 1020.30; ACR: WITH COMPRESSION PADDLE

Tube Voltage (kV)	MQSA—FDA kVp/100	ACR kVp/100 + 0.03
24	0.24	0.27
26	0.26	0.29
28	0.28	0.31
30	0.30	0.33
32	0.32	0.35

ACR, American College of Radiology; CFR, Code of Federal Regulations; FDA, Food and Drug Administration; MQSA, Mammography Quality Standards Act.

tube for 25 to 30 kVp, the beam is “harder” than optimal, which is likely caused by a pitted anode or aged tube, and can potentially result in reduced output and poor image quality. ACR accreditation guidelines recommend a *maximum* HVL based on the target/filter combination listed in Table 8-3. Estimation of radiation dose to the breast requires accurate assessment of the HVL (see Radiation Dosimetry, below).

Tube Output

Tube output [milliroentgen/milliampere second (mR/mAs)] is a function of kVp, target, filtration, location in the field, and distance from the source. At mammography energies, the increase in exposure output is roughly proportional to the third power of the kV, and directly proportional to the atomic number of the target (Fig. 8-14). The MQSA requires that systems used after October 2002 be capable of producing an output of at least 7.0 mGy air kerma per second (800 mR/sec) when operating at 28 kVp in the standard (Mo/Mo) mammography mode at any SID where the system is designed to operate.

Collimation

Fixed-size metal apertures or variable field size shutters collimate the x-ray beam. For most mammography examinations, the field size matches the film cassette sizes (e.g., 18 × 24 cm or 24 × 30 cm). The exposure switch is enabled only when the collimator is present. Many new mammography systems have automatic collimation systems that sense the cassette size. Variable x-y shutters on some systems allow

TABLE 8-3. ACR RECOMMENDATIONS FOR MAXIMUM HVL: MAXIMUM HVL (mm Al) = kVp/100 + C FOR TARGET/FILTER COMBINATIONS

Tube Voltage (kV)	Mo/Mo C = 0.12	Mo/Rh C = 0.19	Rh/Rh C = 0.22	W/Rh C = 0.30
24	0.36	0.43	0.46	0.54
26	0.38	0.45	0.48	0.56
28	0.40	0.47	0.50	0.58
30	0.42	0.49	0.52	0.60
32	0.44	0.51	0.54	0.62

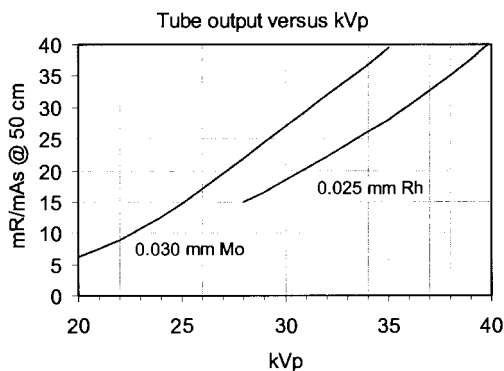


FIGURE 8-14. Tube output (mR/mAs at 50 cm) for a clinical mammography unit with compression paddle in the beam for a Mo target with 0.030-mm Mo filter, and with a 0.025-mm Rh filter. Calculated tube output at 60 cm for a 3-second exposure (100 mA) using 26 kVp (Mo/Mo) is 1,180 mR/sec.

the x-ray field to be more closely matched to the breast volume. Although in principle this seems appropriate, in practice the large unexposed area of the film from tight collimation allows a large fraction of light transmission adjacent to the breast anatomy on a light box, and can result in poor viewing conditions. Collimation to the full area of the cassette is the standard of practice. There is no significant disadvantage to full-field collimation compared to collimation to the breast only, as the tissues are fully in the beam in either case.

The collimator light and mirror assembly visibly define the x-ray beam area. Between the collimator and tube port is a low attenuation mirror that reflects light from the collimator lamp. Similar to conventional x-ray equipment, the light field must be congruent with the actual x-ray field to within 1% of the SID for any field edge, and 2% overall. The useful x-ray field must extend to the chest wall edge of the image receptor without field cutoff. If the image shows collimation of the x-ray field on the chest-wall side or if the chest wall edge of the compression paddle is visible in the image, the machine must be adjusted or repaired.

8.2 X-RAY GENERATOR AND PHOTOTIMER SYSTEM

A dedicated mammography x-ray generator is similar to a standard x-ray generator in design and function. Minor differences include the voltage supplied to the x-ray tube, space charge compensation requirements, and automatic exposure control circuitry.

X-Ray Generator and Automatic Exposure Control

High-frequency generators are the standard for mammography due to reduced-voltage ripple, fast response, easy calibration, long-term stability, and compact size. Older single-phase and three-phase systems provide adequate capabilities for mammography, but certainly without the accuracy and reproducibility of the high-frequency generator (see Chapter 5 for details).

The automatic exposure control (AEC), also called a phototimer, employs a radiation sensor (or sensors), an amplifier, and a voltage comparator, to control the exposure (Fig. 8-15). Unlike most conventional x-ray machines, the AEC detector is located *underneath* the cassette. This sensor consists of a single ionization cham-

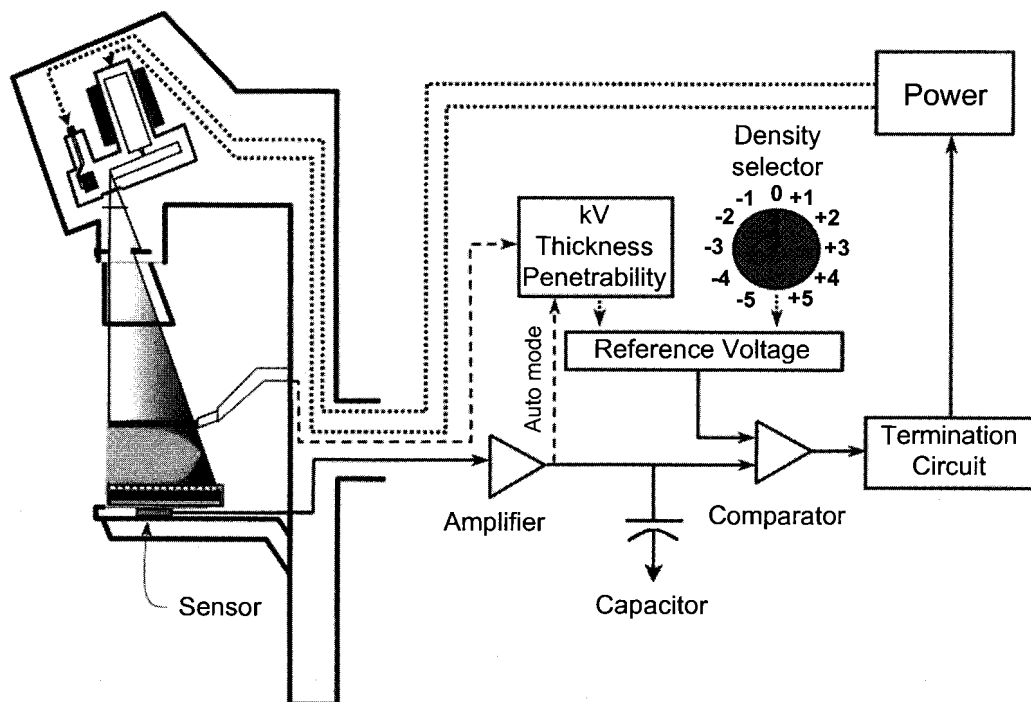


FIGURE 8-15. The automatic exposure control (AEC) (phototimer) circuits use sophisticated algorithms with input from the breast thickness (from the compression paddle), breast density (via a test exposure), density setting (operator density selector), and exposure duration (reciprocity law failure issues) to determine overall exposure time. In fully automatic modes, the kVp, beam filter, and target are also determined, usually from a short (100 msec) exposure that the sensor measures to determine attenuation characteristics of the breast.

ber or an array of three or more semiconductor diodes. The sensor measures the residual x-ray photon flux transmitted through the breast, antiscatter grid (if present), and the image receptor. During the exposure, x-ray interactions in the sensor release electrons that are collected and charge a capacitor. When the voltage across the capacitor matches a preset reference voltage in a comparator switch, the exposure is terminated.

Phototimer algorithms take into account the radiographic technique (kVp, target/filter) and, in the case of extended exposure times, reciprocity law failure (explained in Chapter 6), to achieve the desired film optical density. In newer systems, the operator has several options for exposure control. A fully automatic AEC sets the optimal kV and filtration (and even target material on some systems) from a short test exposure of ~100 msec to determine the penetrability of the breast.

A properly exposed clinical mammogram image should have an optical density of at least 1.0 in the densest glandular regions in the film image. (Note: the mammography phantom, described later, is typically used for calibration of the AEC; a phantom image should have a background optical density between 1.5 and 2.0). At the generator control panel, a density adjustment selector can modify the AEC response. Usually, ten steps (–5 to +5) are available to increase or decrease exposure by adjusting the reference voltage provided to of the AEC comparator switch. A dif-

ference of 10% to 15% positive (greater exposure) or negative (lesser exposure) per step from the neutral setting permits flexibility for unusual imaging circumstances, or allows variation for radiologist preference in film optical density. If the transmission of photons is insufficient to trigger the comparator switch after an extended exposure time, a *backup timer* terminates the exposure. For a retake, the operator must select a *higher* kVp for greater beam penetrability and shorter exposure time.

Inaccurate phototimer response, resulting in an under- or overexposed film, can be caused by breast tissue composition heterogeneity (adipose versus glandular), compressed thickness beyond the calibration range (too thin or too thick), a defective cassette, a faulty phototimer detector, or an inappropriate kVp setting. Film response to very long exposure times, chiefly in magnification studies (low mA capacity of the small focal spot), results in *reciprocity law failure* and inadequate film optical density. For extremely thin or fatty breasts, the phototimer circuit and x-ray generator response can be too slow in terminating the exposure, causing film overexposure.

The position of the phototimer detector (e.g., under dense or fatty breast tissue) can have a significant effect on film density. Previous mammograms can aid in the proper position of the phototimer detector to achieve optimal film density for glandular areas of the breast. Most systems allow positioning in the chest wall to nipple direction, while some newer systems also allow side-to-side positioning to provide flexibility for unusual circumstances. The size of the AEC detector is an important consideration. A small detector (e.g., solid-state device) can be affected by nonuniformities that are not representative of the majority of breast tissue. On the other hand, a large area detector can be affected by unimportant tissues outside the area of interest. Either situation can result in over- or underexposure of the breast image.

Technique Chart

Technique charts are useful guides to determine the appropriate kVp for specific imaging tasks, based on breast thickness and breast composition (fatty versus glandular tissue fractions). Most mammographic techniques use phototiming, and the proper choice of the kVp is essential for a reasonable exposure time, defined as a range from approximately 0.5 to 2.0 seconds, to achieve an optical density of 1.5 to 2.0. Too short an exposure can cause visible grid lines to appear on the image, while too long an exposure can result in breast motion, either of which degrades the quality of the image. Table 8-4 lists kVp recommendations, determined from computer

TABLE 8-4. RECOMMENDED kVp TECHNIQUES, AS A FUNCTION OF BREAST COMPOSITION AND THICKNESS IS DETERMINED WITH COMPUTER SIMULATIONS AND CLINICAL MEASUREMENT^a

Breast Composition	Breast Thickness (cm)						
	2	3	4	5	6	7	8
Fatty	24	24	24	24	25	27	30
50/50	24	24	24	25	28	30	32
Glandular	24	24	26	28	31	33	35

^aThe desired end point is an exposure time between 0.5 and 2.0 sec. This is for a Mo target and 0.030-mm Mo filter using a 100-mA tube current.

simulations and experimental measurements, to meet the desired exposure time range for breast thickness and breast composition.

8.3 COMPRESSION, SCATTERED RADIATION, AND MAGNIFICATION

Compression

Breast compression is a necessary part of the mammography examination. Firm compression reduces overlapping anatomy and decreases tissue thickness of the breast (Fig. 8-16A). This results in fewer scattered x-rays, less geometric blurring of anatomic structures, and lower radiation dose to the breast tissues. Achieving a uniform breast thickness lessens exposure dynamic range and allows the use of higher contrast film.

Compression is achieved with a compression paddle, a flat Lexan plate attached to a pneumatic or mechanical assembly (Fig. 8-16B). The compression paddle should match the size of the image receptor (18×24 cm or 24×30 cm), be flat and parallel to the breast support table, and not deflect from a plane parallel to the receptor stand by more than 1.0 cm at any location when compression is applied. A right-angle edge at the chest wall produces a flat, uniform breast thickness when compressed with a force of 10 to 20 newtons (22 to 44 pounds). One exception is the spot compression exam. A smaller compression paddle area (~5 cm diameter) reduces even further the breast thickness in a specific breast area and redistributes the breast tissue for improved contrast and anatomic rendition, as illustrated in Fig. 8-16C. Spot compression is extremely valuable in delineating anatomy and achieving minimum thickness in an area of the breast presenting with suspicious findings on previous images.

Typically, a hands-free (e.g., foot-pedal operated), power-driven device, operable from both sides of the patient, adjusts the compression paddle. In addition, a mechanical adjustment control near the paddle holder allows fine adjustments of compression. While firm compression is not comfortable for the patient, it is often the difference between a clinically acceptable and unacceptable image.

Scattered Radiation and Antiscatter Grids

X-rays transmitted through the breast contain primary and scattered radiation. Primary radiation retains the information regarding the attenuation characteristics of the breast and delivers the maximum possible subject contrast. Scattered radiation is an additive, slowly varying radiation distribution that degrades subject contrast. If the maximum subject contrast without scatter is $C_0 = \Delta P/P$, the maximum contrast with scatter is

$$C_s = C_0 \left(1 + \frac{S}{P} \right)^{-1}$$

where S is the amount of scatter, P is the amount of primary radiation, and S/P is the scatter to primary ratio. The quantity modifying C_0 is the *scatter degradation factor* or *contrast reduction factor* as explained in Chapter 6. The amount of scatter in mammography increases with increasing breast thickness and breast area, and is relatively constant with kVp. Typical S/P ratios are plotted in Fig. 8-17. For a typical

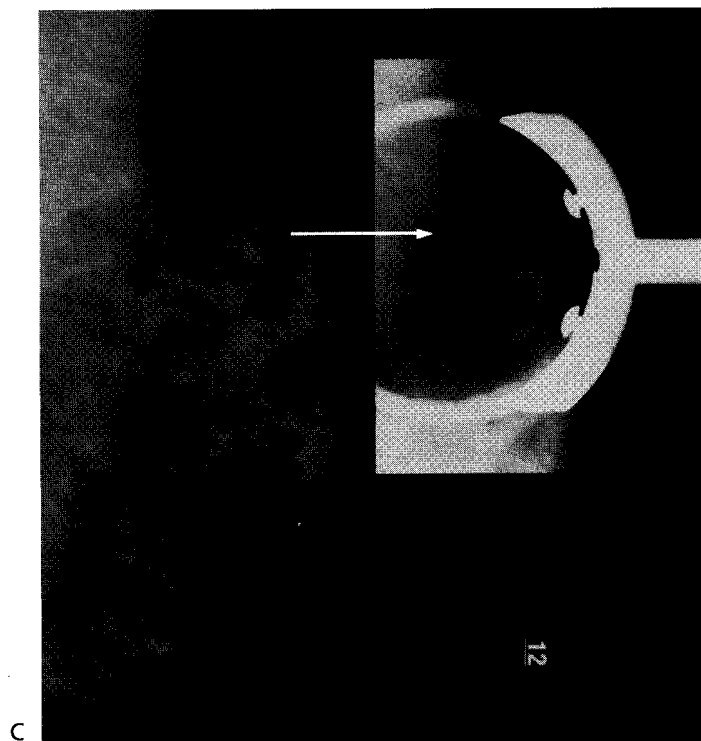
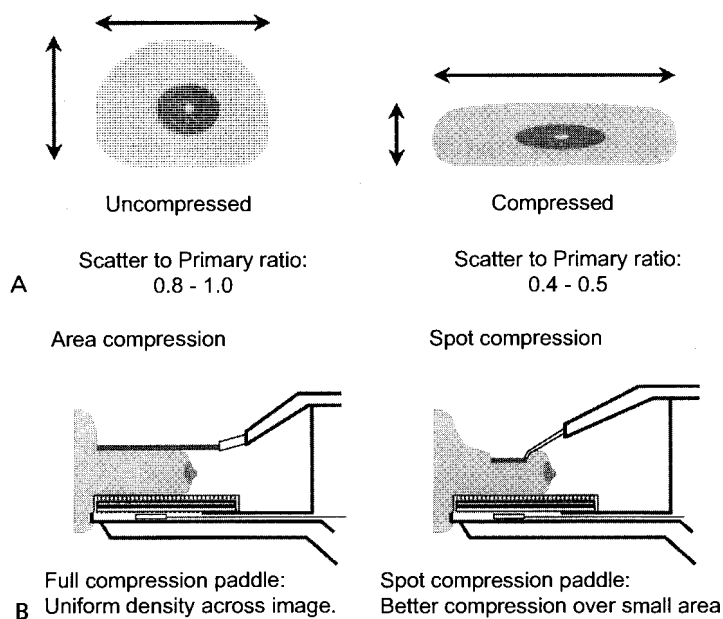


FIGURE 8-16. A: Compression is essential for mammographic studies to reduce breast thickness (lower scatter, reduced radiation dose, and shorter exposure time) and to spread out superimposed anatomy. **B:** Suspicious areas often require "spot" compression to eliminate superimposed anatomy by further spreading the breast tissues over a localized area. **C:** Example of image with suspicious finding (left). Corresponding spot compression image shows no visible mass (right).

6-cm-thick breast, the S/P value is ~ 0.6 and the corresponding scatter degradation factor is $1/1.6 = 0.625 = 62.5\%$. Without some form of scatter rejection, therefore, only 50% to 70% of the inherent subject contrast can be detected. The fraction of scattered radiation in the image can be greatly reduced by the use of *antiscatter grids* or *air gaps*, as well as vigorous compression.

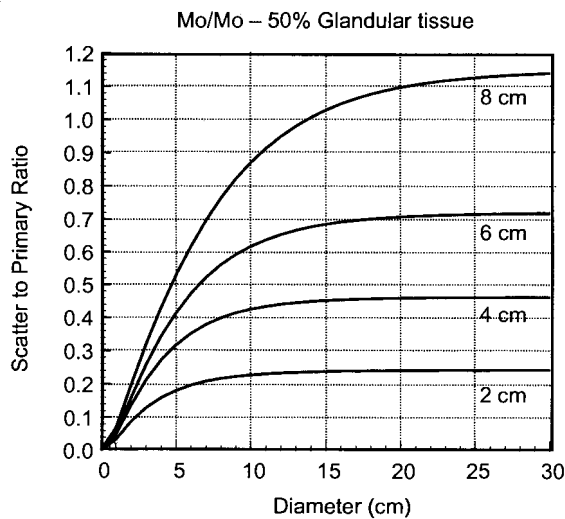


FIGURE 8-17. X-ray scatter reduces the radiographic contrast of the breast image. Scatter is chiefly dependent on breast thickness and field area, and largely independent of kVp in the mammography energy range (25 to 35 kVp). The scatter-to-primary ratio is plotted as a function of the diameter of a semicircular field area aligned to the chest wall edge, for several breast thicknesses of 50% glandular tissue.

Scatter rejection is accomplished with an antiscatter grid for contact mammography (Fig. 8-18). The grid is placed between the breast and the image receptor. Parallel linear grids with a grid ratio of 4:1 to 5:1 are commonly used. Higher grid ratios provide greater x-ray scatter removal but also a greater dose penalty. Aluminum and carbon fiber are typical interspace materials. Grid frequencies (lead strip densities) range from 30 to 50 lines/cm for moving grids, and up to 80 lines/cm for stationary grids. The moving grid oscillates over a distance of approximately 20 grid lines during the exposure to blur the grid line pattern. Short exposures often cause grid-line artifacts that result from insufficient motion. A cellular grid structure designed by one manufacturer provides scatter rejection in two dimensions. The two-dimensional grid structure requires a specific stroke and travel distance during the exposure for complete blurring, so specific exposure time increments are necessary.

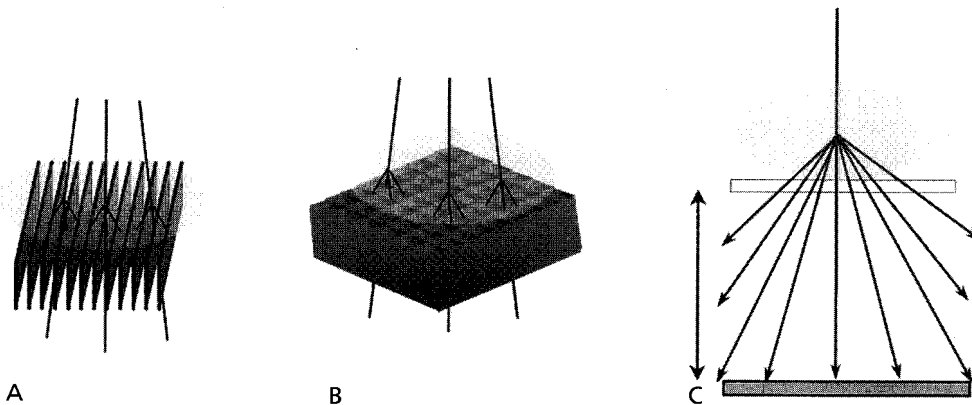


FIGURE 8-18. Antiscatter devices commonly employed in mammography include (A) the linear grid of typically 5:1 grid ratio, (B) a cellular grid structure that rejects scatter in two dimensions, and (C) the air gap that is intrinsic to the magnification procedure.

Grids impose a dose penalty to compensate for x-ray scatter and primary radiation losses that otherwise contribute to film density. The *Bucky factor* is the ratio of exposure with the grid compared to the exposure without the grid for the same film optical density. For mammography grids, the Bucky factor is about 2 to 3, so the breast dose is doubled or tripled, but image contrast improves up to 40%. Grid performance is far from ideal, and optimization attempts have met with mixed success. Future digital slot-scan acquisition methods promise effective scatter reduction without a significant dose penalty (see Chapter 11 on digital mammography).

Air gap techniques minimize scatter by reducing the solid angle subtended by the receptor—in essence, a large fraction of scattered radiation misses the detector. The breast is positioned further from the image receptor and closer to the focal spot. Since the grid is not necessary, the Bucky factor dose penalty is eliminated; however, reduction of the breast dose is offset by the shorter focal spot to skin distance.

Magnification Techniques

The use of geometric magnification is often helpful and sometimes crucial for diagnostic mammography exams. Magnification is achieved by placing a breast support platform (a magnification stand) to a fixed position above the detector, selecting the small (0.1 mm) focal spot, replacing the antiscatter grid with a cassette holder, and using an appropriate compression paddle (Fig. 8-19). Most dedicated mammographic units offer magnification of 1.5 $\times$, 1.8 $\times$, or 2.0 $\times$. Advantages of magnification include (a) increased effective resolution of the image receptor by the magnification factor, (b) reduction of effective image noise, and (c) reduction of scattered radiation. The enlargement of the x-ray distribution pattern relative to the inherent unsharpness of the image receptor provides better effective resolution. However, magnification requires a focal spot small enough to limit geometric blurring, most often a 0.1-mm nominal focal spot size. Quantum noise in a magnified image is less compared to the standard contact image, even though the receptor noise remains the same, because there are more x-ray photons *per object area* creating the enlarged image. The air gap between the breast support surface of the magnification stand and the image receptor reduces scattered radiation, thereby improving image contrast and rendering the use of a grid unnecessary.

Magnification has several limitations, the most dominant being geometric blurring caused by the finite focal spot size (Fig. 8-19). Spatial resolution is poorest on the cathode side of the x-ray field (toward the chest wall), where the effective focal spot size is largest. The modulation transfer function (MTF) is a plot of signal modulation versus spatial frequency (see Chapter 10), which describes the signal transfer characteristics of the imaging system components. A point-source focal spot transfers 100% of the signal modulation of an object at all spatial frequencies (a horizontal MTF curve at 100% or 1.0). Area focal spots produce geometric blurring of the object on the imaging detector. This blurring is greater with larger focal spot size and higher magnification, which causes a loss of resolution (high spatial frequencies) in the projected image. Focal spot MTF curves illustrate the need for the 0.1-mm (small) focal spot to maintain resolution, particularly for magnification greater than 1.5 $\times$ (Fig. 8-20).

The small focal spot limits the tube current to about 25 mA, and extends exposure times beyond 4 seconds for acquisitions greater than 100 mAs. Even slight breast motion will cause blurring during the long exposure times. Additional radi-

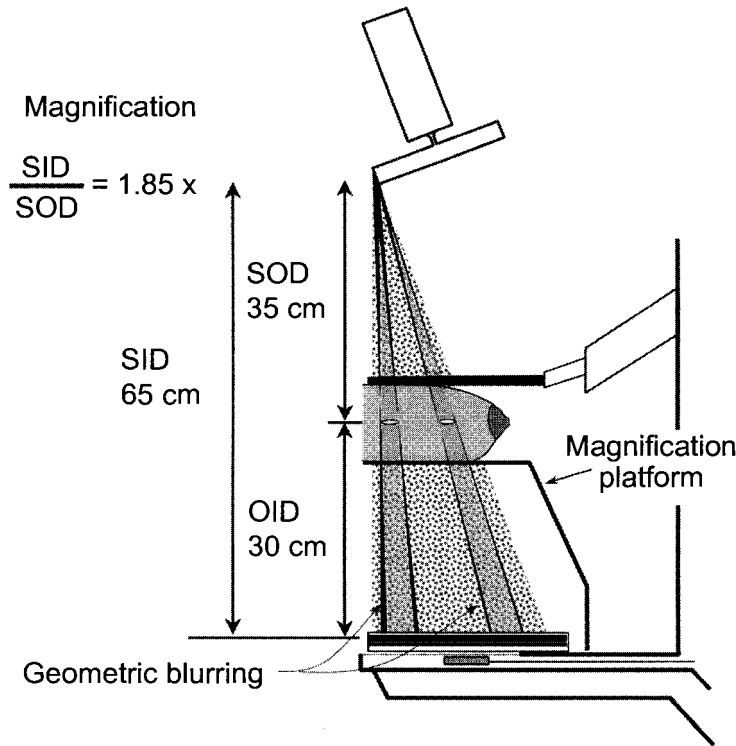


FIGURE 8-19. Geometric magnification. A support platform positions the breast toward the focal spot, giving 1.5× to 2.0× image magnification. A small focal spot (0.1-mm nominal size) reduces geometric blurring and preserves contrast. Variation of focal spot size along the cathode–anode axis is accentuated with magnification due to the increased penumbra width (see Chapter 10) as depicted in the figure, such that the best resolution and detail in the image exists on the anode side of the field toward the nipple.

Focal spot MTF curves

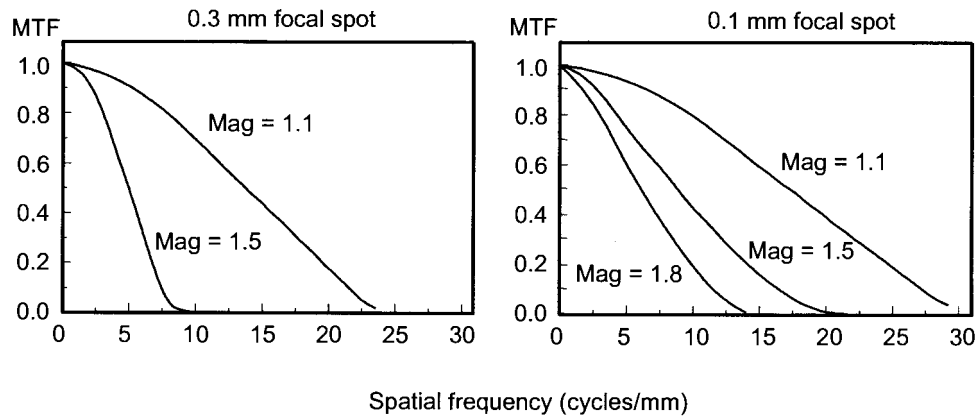


FIGURE 8-20. Modulation transfer function (MTF) curves for 0.3- and 0.1-mm focal spots at various magnifications illustrate the need for a 0.1-mm focal spot in magnification work.

ation exposure is also necessary to compensate for reciprocity law failure. Entrance skin exposure and breast dose are reduced because no grid is used, but are increased due to the shorter focus to breast distance (Fig. 8-19). In general, the average glandular dose with magnification is similar to contact mammography. Higher speed screen-film combinations with lower limiting spatial resolution but higher absorption and/or conversion efficiency are sometimes used for magnification studies to lower radiation dose and reduce exposure time, because the “effective” detector resolution is increased by the magnification factor.

8.4 SCREEN-FILM CASSETTES AND FILM PROCESSING

Screen-film technology has improved dramatically over the past 20 years for mammography, from direct exposure industrial x-ray film, to vacuum packed screen-film systems, to contemporary high-speed and high-contrast screen-film detectors. Film processing, often neglected in the past, is now recognized as a critical component of the imaging chain.

Design of Mammographic Screen-Film Systems

Image receptors must provide adequate spatial resolution, radiographic speed, and image contrast. Screens introduce the classic trade-off between resolution and radiographic speed. Film does not limit spatial resolution, but does significantly influence the contrast of the image. High-speed film can cause an apparent decrease in contrast sensitivity because of increased quantum noise and grain noise.

Mammography cassettes have unique attributes. Most cassettes are made of low-attenuation carbon fiber and have a single high-definition phosphor screen used in conjunction with a single emulsion film. Terbium-activated gadolinium oxysulfide ($\text{Gd}_2\text{O}_2\text{S:Tb}$) is the most commonly used screen phosphor. The green light emission of this screen requires green-sensitive film emulsions. Intensifying screen-film speeds and spatial resolution are determined by phosphor particle size, light absorbing dyes in the phosphor matrix, and phosphor thickness. Mammography screen-film speeds are classified as regular (100 or par speed) and medium (150 to 190 speed), where the 100-speed system requires ~12 to ~15 mR radiation exposure to achieve the desired film optical density. For comparison, a conventional “100-speed” screen film cassette requires about 2 mR.

The screen is positioned in the *back* of the cassette so that x-rays travel through the cassette cover and the film before interacting with the phosphor. Because of exponential attenuation, x-rays are more likely to interact near the phosphor surface closest to the film emulsion. This reduces the distance traveled by the light, minimizing the spreading of the light and thereby preserving spatial resolution. X-ray absorption at depth in the screen produces a broader light distribution, and reduces resolution (Fig. 8-21). For a 100-speed mammography screen-film cassette, the limiting spatial resolution is about 15 to 20 lp/mm.

Screen-Film Systems for Mammography

A partial list of film-screen systems available as of the year 2001 are listed in Table 8-5. Almost all cassettes have a single gadolinium oxysulfide phosphor screen and

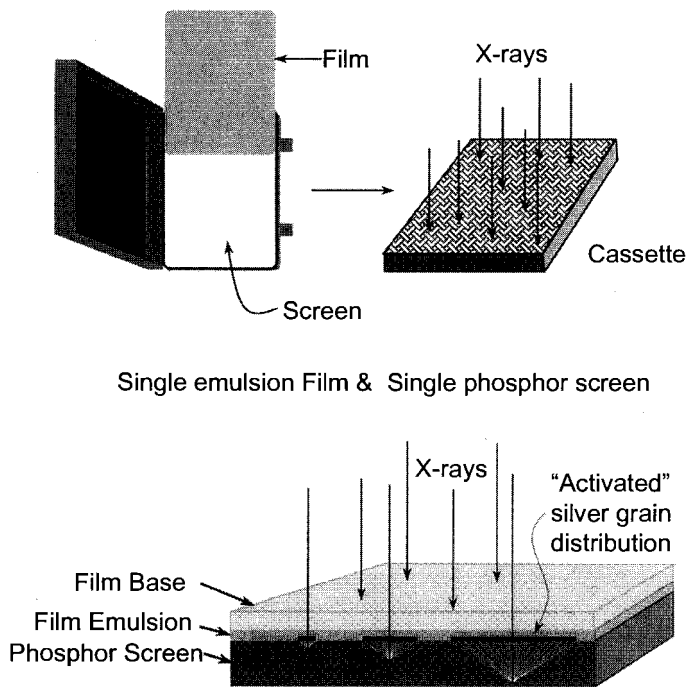


FIGURE 8-21. Screen-film systems used in mammography most often employ a single phosphor screen with a single emulsion film, and are “front-loaded” (x-rays pass through the film to the screen). Light spread varies with the depth of x-ray absorption, as depicted in the lower diagram. Most x-rays interact in the layers of the screen closest to the film, thereby preserving spatial resolution.

therefore require a green-sensitive single emulsion film. In addition, manufacturers have designed the screen and the film to be matched for optimal image quality. It is not recommended to mix screens and films, because of the potential variation in speed and contrast characteristics.

Comparison of Conventional and Mammography Screen-Film Detectors

Figure 8-22 compares characteristic and MTF curves for direct exposure film, mammographic screen-film, and conventional screen-film. Direct exposure films have higher spatial resolution, but require 50 to 100 times more exposure than

TABLE 8-5. MAMMOGRAPHY FILMS AND SCREENS IN CURRENT USE^a

Manufacturer	Par-Speed Screen	Medium-Speed Screen ^b	Film (Matched with Screen)
Agfa	HD	HD-S	Mamoray HDR, HDR-C, HT, Microvision Ci
Fuji	AD fine UM fine	AD medium UM medium	AD-M UM-MA HC
Kodak	MIN-R	MIN-R 2000, MIN-R 2190 MIN-R medium	MIN-R 2000, MIN-R L MIN-R 2000, MIN-R L

Note: Microvision Ci film is also used with the Kodak MIN-R screen.

^aAs of 2000.

^bMedium speed is approximately 30% to 40% faster than the par speed for the same OD.

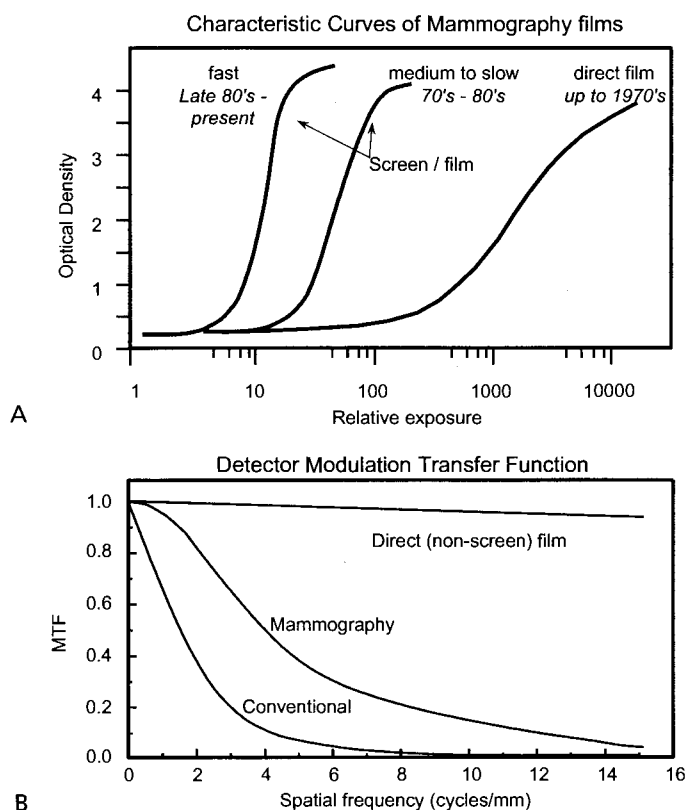


FIGURE 8-22. A: Comparison of characteristic curves (H and D curves) of direct exposure, slow to medium speed, and fast mammography screen-films. **B:** Sample MTF curves indicate the superior transfer characteristics of a mammography detector compared to a conventional film screen detector, and demonstrate an inverse relationship between screen-film speed and spatial resolution.

mammographic screen-film and have lower film contrast. Mammographic screen-film has higher resolution and higher film contrast than conventional radiographic screen-film, but also has less exposure latitude, which can be a problem when imaging thick, dense breasts.

The selection of a screen-film combination involves such measures as resolution and radiographic speed, as well as factors such as screen longevity, cassette design, film base color, film contrast characteristics, radiologist preference, and cost. In today's market, there is no single best choice; hence, most screen-film receptors, when used as recommended with particular attention to film processing, provide excellent image quality for mammography imaging.

Film Processing

Film processing is a critical step in the mammographic imaging chain. Consistent film speed, contrast, and optical density levels are readily achieved by following the manufacturer's recommendations for developer formulation, development time, developer temperature, replenishment rate, and processor maintenance. If the developer temperature is too low, the film speed and film contrast will be reduced,

requiring a compensatory increase in radiation dose. If the developer temperature is too high or immersion time too long, an increase in film speed occurs, permitting a lower dose; however, the film contrast is likely to be reduced while film fog and quantum mottle are increased. Stability of the developer may also be compromised at higher-than-recommended temperatures. The manufacturer's guidelines for optimal processor settings and chemistry specifications should be followed.

Film Sensitometry

Because of the high sensitivity of mammography film quality to slight changes in processor performance, routine monitoring of proper film contrast, speed, and base plus fog values is important. A film-processor quality control program is required by MQSA regulations, and daily sensitometric tests *prior to the first clinical images* must verify acceptable performance. A sensitometer, densitometer, thermometer, and monitoring chart are tools necessary to complete the processor evaluation (Fig. 8-23). (Sensitometers and densitometers are discussed in Chapter 7.) A film obtained from a box of film reserved for quality control (QC) testing (the QC film eliminates variations between different lots of film from the manufacturer) is exposed using a calibrated sensitometer with a spectral output similar to the phosphor (e.g., $\text{Gd}_2\text{O}_2\text{S:Tb}$ screens largely emit green light). The processor then develops the film. On the film, the optical densities produced by the calibrated light intensity strip of the sensitometer are measured with a film densitometer. Data points corresponding to the base plus fog optical density, a speed index step around an OD of 1.0, and a density difference (an index of contrast) between two steps of ~ 0.5 and ~ 2.0 OD are plotted on the processor QC chart. Action limits (*horizontal dashed lines* on the graphs shown in Fig. 8-24) define the range of acceptable processor performance. If the test results fall outside of these limits, corrective action must be taken. Developer temperature measurements are also a part of the daily processor quality control testing. Typical processor developer temperatures are $\sim 35^\circ\text{C}$ (95°F). With good daily quality control procedures and record keeping, processor problems can be detected before they harm clinical image quality.

Variation in film sensitivity often occurs with different film lots (a "lot" is a specific manufacturing run). A "crossover" measurement technique must be used before using a new box of film for QC. This requires the sensitometric evaluation of five films from the nearly empty box of film and five films from the new box of film. Individual OD steps on the old and new group of films are measured densitometrically, averages are determined, and a difference is calculated. A new baseline is adjusted from the old baseline by this positive or negative value. Action limits are then established for the new baseline level.

A film characteristic curve and gradient curve (the gradient is the rate of change of the characteristic curve) are shown in Fig. 8-25. The gradient curve shows the film contrast as a function of incident exposure. Sometimes more useful is the plot of the gradient versus film optical density (Fig. 8-26), depicting the contrast response of the film at different levels of optical density. Small changes in the shape of the characteristic curve can lead to relatively large changes in the gradient versus OD curve, making this a sensitive method for monitoring processor and film response over time.

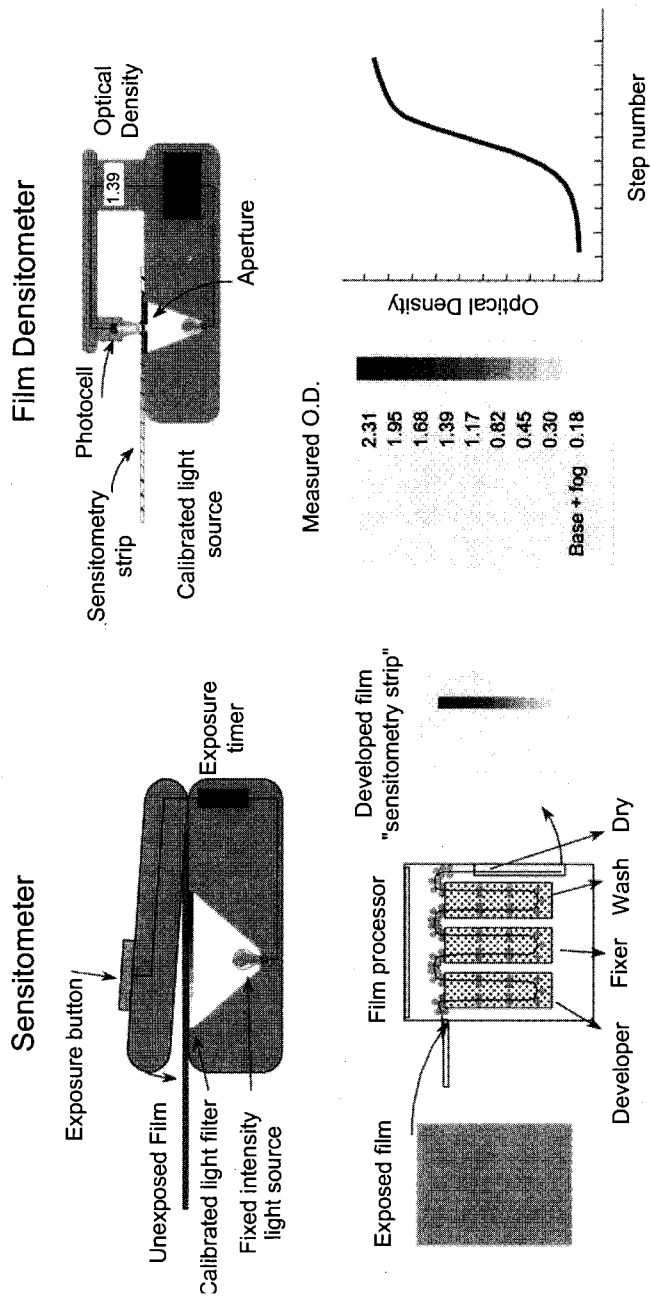


FIGURE 8-23. Film processor testing requires a calibrated sensitometer to "flash" an unexposed film to a specified range of light intensities using an optical step wedge. The film is then developed. The corresponding optical densities on the processed film are measured with a densitometer. A characteristic curve is a graph of film optical density (vertical axis) versus step number (horizontal axis).

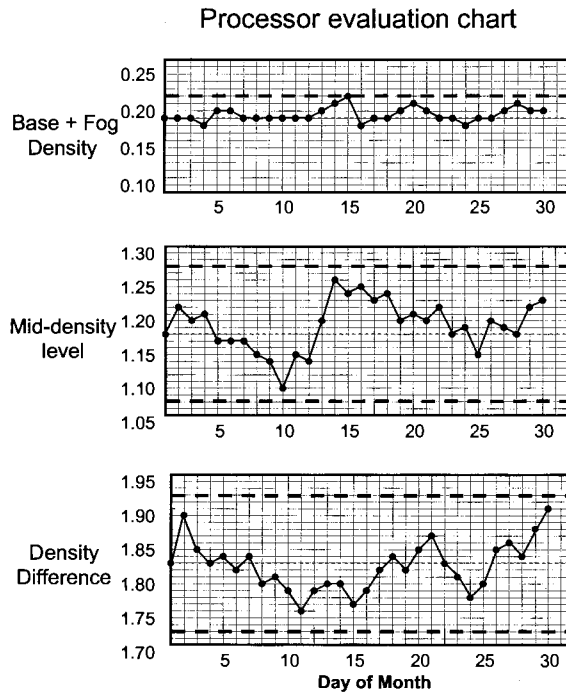


FIGURE 8-24. The processor evaluation chart uses sensitometry data to track automatic film processor performance daily. Control limits are indicated by *dashed lines* on each graph. When values extend beyond the limits, action must be taken to correct the problem before any clinical images are acquired and processed.

Extended Cycle Processing

Extended cycle processing (also known as “push processing”) is a film development method that increases the speed of some single emulsion mammography films by extending the developer immersion time by a factor of two (usually from ~20 to ~40 seconds). The rationale behind extended processing is to allow for the complete development of the small grain emulsion matrix used in high-resolution mammography films, where the standard 22-second immersion time is not sufficient. The extended development time allows further reduction of the latent image centers and deposition

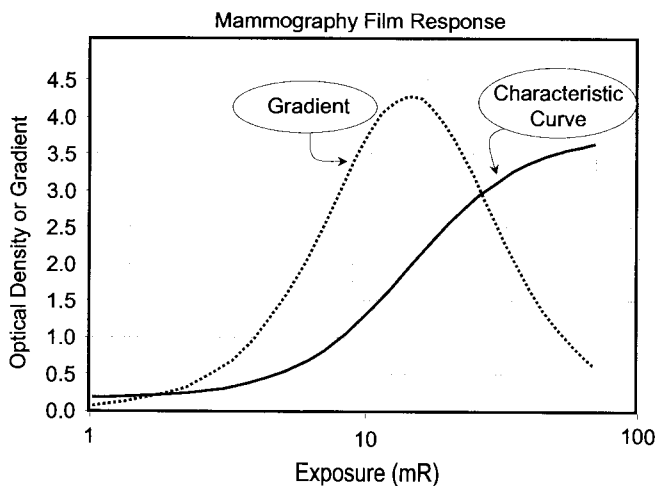


FIGURE 8-25. The characteristic curve is a graph of the optical density as a function of the logarithm of exposure. The gradient curve (the rate of change of the characteristic curve) indicates the contrast as a function of the logarithm of exposure. In this example, the maximal value of the gradient curve is achieved for an exposure of about 15 mR. The characteristic curve shows an optical density of about 2.0 for this exposure. Thus, the maximal contrast for this film-screen system is achieved for an optical density of about 2.0.

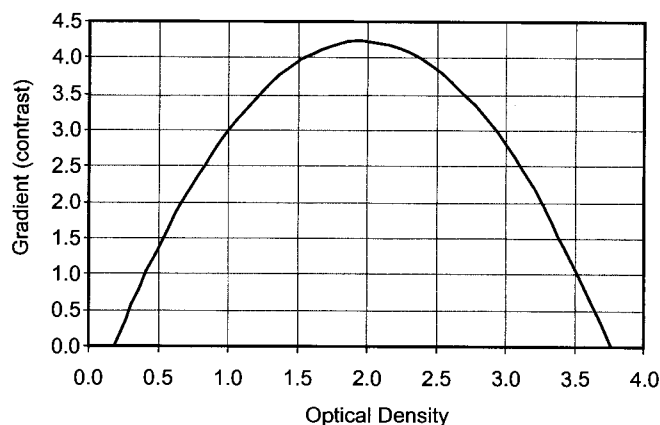


FIGURE 8-26. A plot of film gradient versus optical density shows the film optical density that generates the highest contrast.

of silver grains that would otherwise be removed in the fixer solution. For many single emulsion films (depending on the manufacturer and type of film), extended developer immersion time increases the film contrast and speed, providing up to a 35% to 40% decrease in the required x-ray exposure compared to standard processing for the same film optical density. The disadvantages are reduced film throughput and less forgiving processing conditions. Extended processing on a conventional 90-second processor requires slowing the transport mechanism, which doubles the processing time and decreases film throughput. Dedicated processors with extended processing options increase the developer immersion time only, without changing either the fix or wash cycle times, for an increase of only 20 to 30 seconds per film.

In response to extended processing, manufacturers have developed films that provide the benefits of extended cycle processing with standard processing. In fact, extended processing is becoming a technique of the past. Figure 8-27 shows the

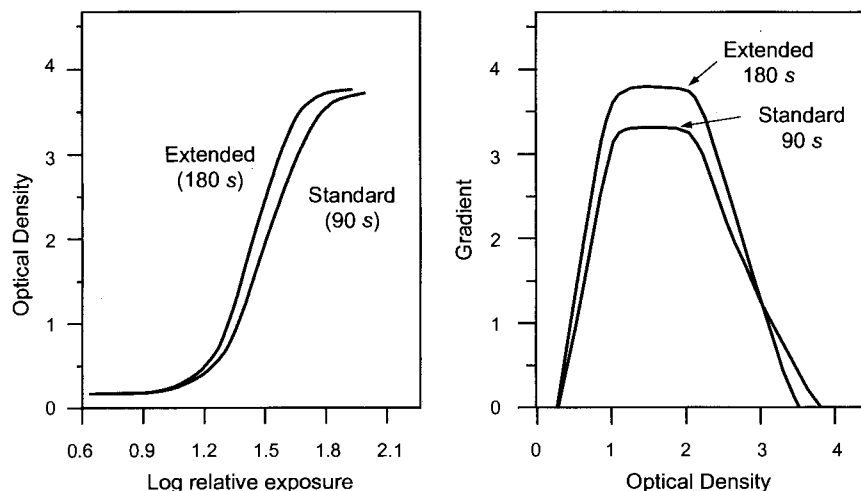


FIGURE 8-27. Response of film to normal and extended (180-second) processing conditions shows an advantage in speed and contrast for the extended cycle. The gradient versus OD curve (*right*) more clearly shows the differences in the contrast response of the two films.

characteristic curves, the corresponding gradient versus optical density curves, and the sensitivity of the gradient versus OD relationship for standard and extended processing.

Film Viewing Conditions

Subtle objects recorded on the mammographic film can easily be missed under sub-optimal viewing conditions. Since mammography films are exposed to high optical densities to achieve high contrast, view boxes providing a high *luminance* (the physical measure of the brightness of a light source, measured in candela (cd/m^2)) are necessary. (For example, the luminance of $1 \text{ cd}/\text{m}^2$ is equivalent to the amount of luminous intensity of a clear day, $\sim 1/4$ hour after sunset.) The luminance of a mammography view box should be at least $3,000 \text{ cd}/\text{m}^2$. In comparison, the luminance of a typical view box in diagnostic radiology is about $1,500 \text{ cd}/\text{m}^2$.

Film masking (blocking clear portions of the film and view box) is essential for preserving perceived radiographic contrast. The film should be exposed as close as possible to its edges and the view box should have adjustable shutters for masking.

The ambient light intensity in a mammography reading room should be low to eliminate reflections from the film and to improve perceived radiographic contrast. *Illuminance* (the luminous flux incident upon a surface per unit area, measured in lux or lumens/ m^2) should be at very low levels (e.g., below 50 lux). (The amount of illumination from a full moon is about 1 lux, and normal room lighting ranges from 100 to 1,000 lux.) Subdued room lighting is necessary to reduce illuminance to acceptable levels.

A high-intensity “bright” light should be available to penetrate high optical density regions of the film, such as the skin line and the nipple area. In addition, use of a magnifying glass aids the visibility of fine detail in the image, such as microcalcifications.

8.5 ANCILLARY PROCEDURES

Stereotactic Breast Biopsy

Stereotactic breast biopsy systems provide the capability to localize in three dimensions and physically sample suspected lesions found on mammograms. Two views of the breast are acquired from different x-ray tube angles (usually $+15$ and -15 degrees from the normal). The projections of the lesions on the detector shift in the opposite direction of tube motion, dependent on the lesion depth. Figure 8-28 illustrates the resultant stereotactic images and the relationship between the shifts of the lesion compared to the lesion depth. Measurement of the shift of a specific lesion permits the calculation of its distance from the detector using trigonometric relationships and enables the determination of the trajectory and depth to insert a biopsy needle. A biopsy needle gun is then positioned and activated to capture the tissue sample.

Early biopsy systems were film-based, requiring extra time for film processing and calculation of needle coordinates and trajectory. In the early 1990s, dedicated prone digital imaging systems were introduced to automate the calculations and provide immediate image acquisition for fast and accurate biopsy procedures. The image receptors of these systems are based on charge-coupled-device (CCD) cam-

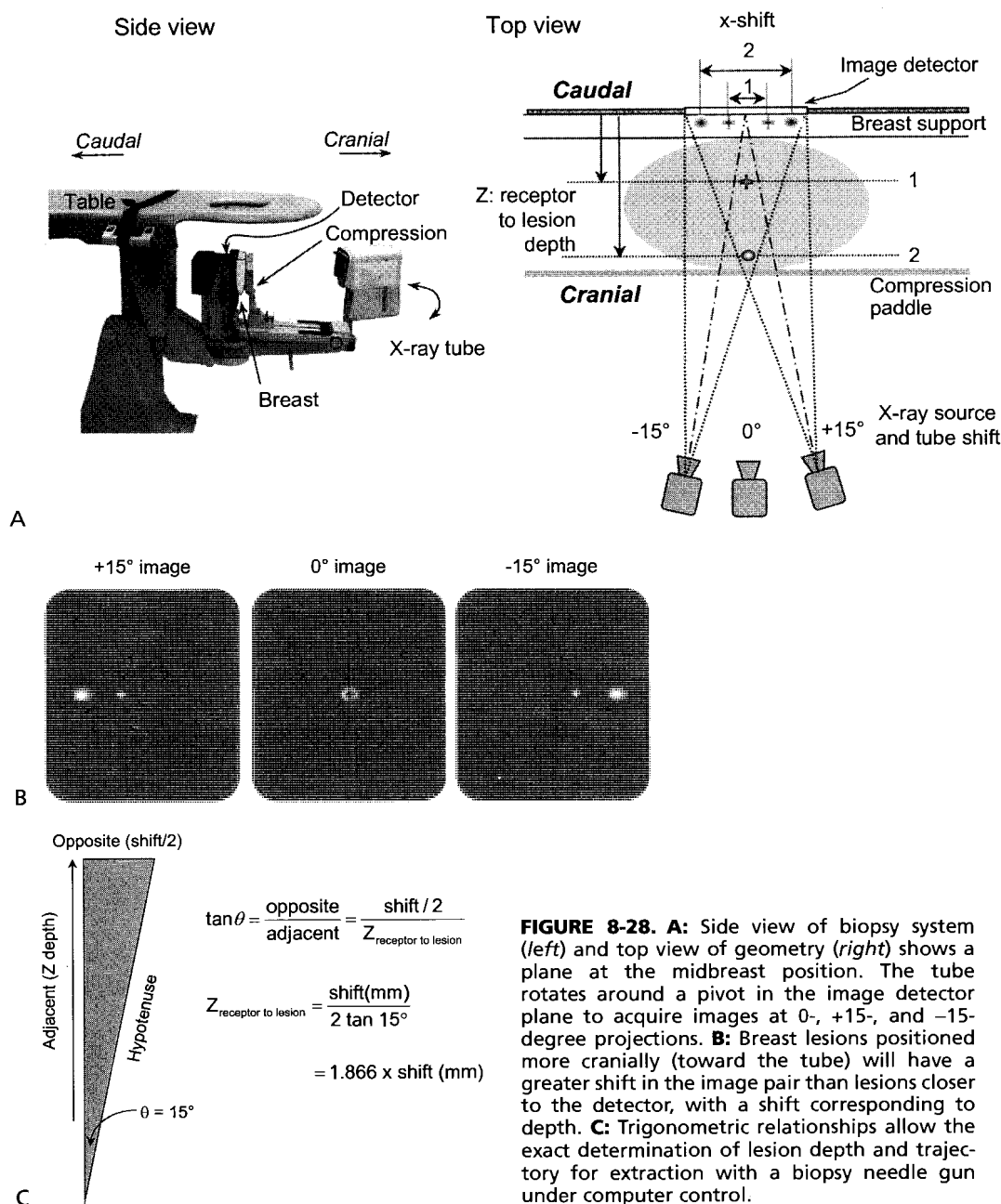


FIGURE 8-28. A: Side view of biopsy system (left) and top view of geometry (right) shows a plane at the midbreast position. The tube rotates around a pivot in the image detector plane to acquire images at 0-, +15-, and -15-degree projections. **B:** Breast lesions positioned more cranially (toward the tube) will have a greater shift in the image pair than lesions closer to the detector, with a shift corresponding to depth. **C:** Trigonometric relationships allow the exact determination of lesion depth and trajectory for extraction with a biopsy needle gun under computer control.

eras coupled to x-ray phosphor screens. A CCD is a light-sensitive electronic device that stores electrical charges proportional to the incident light intensity on discrete pixel areas across its sensitive area (see Chapter 11). The pixel dimension is small, on the order of $25 \mu\text{m}$, so a CCD chip of $25 \times 25 \text{ mm}$ has $1,000 \times 1,000$ pixels. To achieve a $5 \times 5 \text{ cm}$ field of view (FOV) requires a demagnification of the image plane by a factor of 2. Optical lenses or fiberoptic tapers are usually employed, thus

achieving a pixel size at the image plane of 50 μm , equivalent to 10 lp/mm. Image acquisition proceeds by integrating x-ray exposure (light) on the CCD photodetector, reading the pixel array one line at a time after the exposure, amplifying the electronic signal, and digitizing the output.

Digital add-on biopsy units with larger FOVs (e.g., 8×8 cm) have recently been introduced using CCDs or other photosensitive detectors that fit into the cassette assemblies of the standard mammography units.

Full-Field Digital Mammography

The next frontier in mammography is the use of full-field-of-view digital receptors to capture the x-ray fluence transmitted through the breast, to optimize post-processing of the images, and to permit computer-aided detection. While full implementation of digital mammography is still some time in the future for screening mammography (mainly because of cost), the potential for improvement in patient care has been demonstrated by small FOV digital biopsy systems. The new millennium has witnessed the introduction of the first Food and Drug Administration (FDA)-approved full-field digital mammography systems. Technologic advances continue in four areas of development: slot-scan geometry using CCD arrays, large area detector using fiberoptic taper CCD cameras, photostimulable phosphor (computed) radiography, and flat-panel detectors using thin film transistor arrays. Refer to Chapter 11 for more details on large area digital detector devices.

There are several advantages to digital mammography. The most important are overcoming some of the limitations of screen-film detectors, improving the conspicuity of lesions, and using computer-aided detection (CAD) to provide a second opinion. Screen-film detectors for mammography are excellent image capture and storage media, but their "Achilles heel" is the limited exposure dynamic range of film (Fig. 8-25 indicates an exposure dynamic range of about 25:1), which is necessary to achieve high radiographic contrast. A highly glandular breast can have an exposure latitude that exceeds 200:1, producing under- or overexposed areas on the film. Digital detectors have dynamic ranges exceeding 1,000:1, and image processing can render high radiographic contrast over all exposure levels in the image.

Technical Requirements for Digital Receptors and Displays in Mammography

Digital detectors are characterized by pixel dimension (which determines spatial sampling), and pixel bit depth (which determines the gray-scale range). In mammography, screen-film detectors have an effective sampling frequency of ~ 20 lp/mm, which is equivalent to a pixel size of $1/(2 \times 20/\text{mm}) = 1/(40/\text{mm}) = 25 \mu\text{m}$. To cover an 18×24 cm detector with the sampling pitch (spacing between samples) equal to the sampling aperture (pixel area) requires a matrix size of $7,200 \times 9,600$ pixels, and even more for the 24×30 cm detector. Considering a breast of 8-cm thickness, the exposure under the dense areas is likely about 0.7 mR, and perhaps 100 times that amount in the least attenuated areas at the skin line, or about 70 mR. A digital gray-scale range of about 700 to 7,000 is therefore necessary to detect a 1% contrast change generated by the exposure differences, requiring the use of 12 bits (providing 4,096 gray levels) to 13 bits (providing 8,192 gray levels) per pixel. A single digital image thus contains $7,200 \times 9,600 \times 13$ bits/pixel, or an uncompressed storage requirement of over 138

megabytes, and four times that for a four-image screening exam. Research studies indicate the possibility of relaxing digital acquisition to 50 μm and even further to 100 μm pixel size. (The FDA has approved a full-field digital mammography device with a 100 μm pixel aperture in early 2000.) Digital systems outperform screen-film receptors in contrast detection experiments.

Disadvantages of digital mammography include image display and system cost. Soft copy (video) displays currently lack the luminance and spatial resolution to display an entire mammogram at full fidelity. Films produced by laser cameras are used for interpretation. The extremely high cost of full-field digital mammography systems (currently about five times more expensive than a comparable screen-film system) hinders their widespread implementation.

8.6 RADIATION DOSIMETRY

X-ray mammography is the technique of choice for detecting nonpalpable breast cancers. However, the risk of carcinogenesis from the radiation dose to the breast is of much concern, particularly in screening examinations, because of the large number of women receiving the exams. Thus, the monitoring of dose is important and is required yearly by the MQSA.

The glandular tissue is most always the site of carcinogenesis, and thus the preferred dose index is the *average glandular dose*. Because the glandular tissues receive varying doses depending on their depths from the skin entrance site of the x-ray beam, estimating the dose is not trivial. The *midbreast dose*, the dose delivered to the plane of tissue in the middle of the breast, was the radiation dosimetry benchmark until the late 1970s. The midbreast dose is typically lower than the average glandular dose and does not account for variation in breast tissue composition.

Average Glandular Dose

The average glandular dose, D_g , is calculated from the following equation:

$$D_g = D_{gN} \times X_{ESE}$$

where X_{ESE} is the entrance skin exposure (ESE) in roentgens, and D_{gN} is an ESE to average glandular dose conversion factor with units of mGy/R or mrad/R. An air-filled ionization chamber measures the ESE for a given kVp, mAs, and beam quality.

The conversion factor D_{gN} is determined by experimental and computer simulation methods and depends on radiation quality (kVp and HVL), x-ray tube target material, filter material, breast thickness, and tissue composition. Table 8-6 lists D_{gN} values for a 50% adipose, 50% glandular breast tissue composition of 4.5-cm thickness as a function of HVL and kVp for a Mo/Mo target/filter combination. For a 26-kVp technique with a HVL of 0.35 mm Al, the average glandular dose is approximately 17% of the measured ESE (Table 8-6). For higher average x-ray energies (e.g., Mo/Rh, Rh/Rh, W/Mo, W/Rh), conversion tables specific to the generated x-ray spectrum must be used, as the D_{gN} values increase due to the higher effective energy of the beam. D_{gN} decreases with an increase in breast thickness for constant beam quality and breast composition. This is because the beam is rapidly attenuated and the glandular tissues furthest from the entrance receive much less dose in the thicker breast (e.g., $D_{gN} = 220$ mrad/R for 3-cm thickness versus 110

TABLE 8-6. D_{gN} CONVERSION FACTOR (mRAD PER ROENTGEN) AS A FUNCTION OF HVL AND kVp FOR Mo TARGET/FILTER: 4.5-CM BREAST THICKNESS OF 50% GLANDULAR AND 50% ADIPOSE BREAST TISSUE COMPOSITION*

HVL (mm)	kVp							
	25	26	27	28	29	30	31	32
0.25	122							
0.26	126	128						
0.27	130	132	134					
0.28	134	136	138	139				
0.29	139	141	142	143	144			
0.30	143	145	146	147	148	149		
0.31	147	149	150	151	152	153	154	
0.32	151	153	154	155	156	158	159	160
0.33	155	157	158	159	160	162	163	164
0.34	160	161	162	163	164	166	167	168
0.35	164	166	167	168	169	170	171	172
0.36	168	170	171	172	173	174	175	176
0.37		174	175	176	177	178	178	179
0.38			179	180	181	182	182	183
0.39				184	185	186	186	187
0.40					189	190	191	192

*Adapted from ACR QC Manual, 1999.

mrads/R for 6-cm thickness in Table 8-6). However, the lower D_{gN} for the thicker breast is more than offset by a large increase in the entrance exposure necessary to achieve the desired optical density on the film.

Factors Affecting Breast Dose

Many variables affect breast dose. The speed of the screen-film receptor and the film optical density preferred by the radiologist are major factors. Breast thickness and tissue composition strongly affect x-ray absorption. Higher kVp (higher HVL) increases beam penetrability (lower ESE and lower average glandular dose), but decreases inherent subject contrast (Fig. 8-29). Antiscatter grids improve image

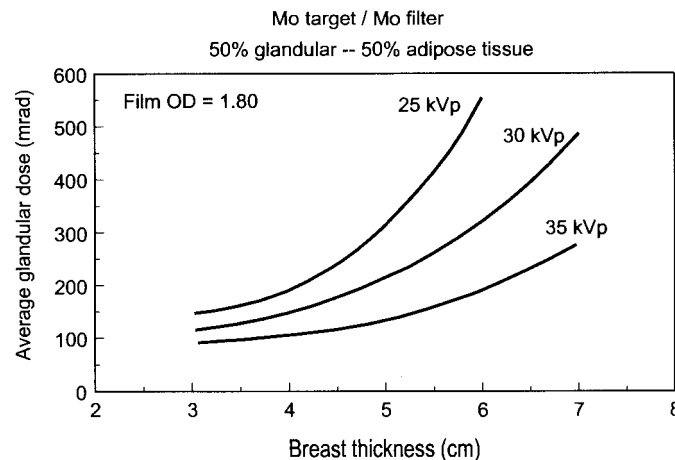


FIGURE 8-29. Approximate average glandular doses (25 to 35 kVp, Mo target/0.03-mm Mo filter) for an optical density of about 1.80 and a 50:50 glandular/adipose tissue content are plotted as a function of breast thickness.

TABLE 8-7. AVERAGE GRANDULAR DOSE IN mRAD WITH THREE TISSUE COMPOSITIONS OF THE BREAST USING kVp VALUES IN TABLE 8.4^a

Breast Composition	Breast Thickness (cm)						
	2	3	4	5	6	7	8
Fatty	50	75	125	190	300	420	430
50/50	65	100	180	280	380	480	520
Glandular	100	180	280	400	500	520	650

^aMo/Mo target/filter, 5:1 grid, and film OD = 1.80 (Fuji AD-M film and AD-fine screens).

quality, but increase radiation dose by the Bucky factor ($\sim 2\times$). Table 8-7 lists measured average glandular doses for a contemporary clinical mammography system with a film optical density of 1.8, a Mo/Mo target/filter, and a kVp recommended in Table 8-4.

The MQSA limits the average glandular dose to 3 mGy (300 mrad) per film (e.g., 6 mGy for two films) for a compressed breast thickness of 4.2 cm and a breast composition of 50% glandular and 50% adipose tissue (e.g., the MQSA-approved mammography phantom). If the average glandular dose for this phantom exceeds 3 mGy, mammography may not be performed. The average glandular dose for this phantom is typically 1.5 to 2.2 mGy per view or 3 to 4.4 mGy for two views for a film optical density of 1.5 to 2.0.

Mammography image quality strongly depends on beam energy, quantum mottle, and detected x-ray scatter. Faster screen-film detectors can use a lower dose, but often at the expense of spatial resolution. Digital detectors with structured phosphors (described in Chapter 11) can improve quantum detection efficiency without reducing resolution. Studies indicate the possibility of improving scatter rejection with highly efficient moving multiple slit designs or air-interspaced grids. Higher effective energy beams (e.g., Rh target and Rh filter) can significantly reduce the dose to thick, dense glandular breasts without loss of image quality. The use of nonionizing imaging modalities such as ultrasound and MRI can assist in reducing the overall radiation dose for a diagnostic breast evaluation.

8.7 REGULATORY REQUIREMENTS

Regulations mandated by the Mammography Quality Standards Act of 1992 specify the operational and technical requirements necessary to perform mammography in the United States. These regulations are contained in Title 21 of the Code of Federal Regulations, Part 900.

Accreditation and Certification

Accreditation and certification are two separate processes. For a facility to perform mammography legally under MQSA, it must be certified and accredited. To begin the process, it must first apply for accreditation from an accreditation body (the American College of Radiology or currently the states of Arkansas, California, Iowa, or Texas, if the facility is located in one of those states). The accreditation body ver-

ifies that the mammography facility meets standards set forth by the MQSA to ensure safe, reliable, and accurate mammography. These include the initial qualifications and continuing experience and education of interpreting physicians, technologists, and physicists involved in the mammography facility. These also include the equipment, quality control program, and image quality. *Certification* is the approval of a facility by the U.S. FDA to provide mammography services, and is granted when accreditation is achieved.

Specific Technical Requirements

Tables 8-8 and 8-9 summarize the annual QC tests performed by the technologist and physicist, respectively, as required by MQSA regulations (1999).

The Mammography Phantom

The mammography phantom is a test object that simulates the radiographic characteristics of compressed breast tissues, and contains components that model breast disease and cancer in the phantom image. Its role is to determine the adequacy of the overall imaging system (including film processing) in terms of detection of subtle radiographic findings, and to assess the reproducibility of image characteristics (e.g., contrast and optical density) over time. It is composed of an acrylic block, a wax insert, and an acrylic disk (4 mm thick, 10 mm diameter) attached to the top of the phantom. It is intended to mimic the attenuation characteristics of a “standard breast” of 4.2-cm compressed breast thickness of 50% adipose and 50% glandular tissue composition. The wax insert contains six cylindrical nylon fibers of decreasing diameter, five simulated calcification groups (Al_2O_3 specks) of decreasing size, and five low contrast disks, of decreasing diameter and thickness, that simulate masses (Fig. 8-30). Identification of the smallest objects of each type that are visible in the phantom image indicates system performance. To pass the MQSA image quality standards, at least four fibers, three calcification groups, and three masses must be clearly visible (with no obvious artifacts) at an average glandular dose of less than 3 mGy. Furthermore, optical density at the center of the phantom image must be at least 1.2 (the ACR recommends at least 1.4) and must not change by more than ± 0.20 from the normal operating level. The optical density difference between the background of the phantom and the added acrylic disk used to assess image contrast must not vary by more than ± 0.05 from the established operating level.

Quality Assurance

Mammography is a very demanding procedure, from both an equipment capability and technical competence standpoint. Achieving the full potential of the examination requires careful optimization of technique and equipment. Even a small change in equipment performance, film processing, patient setup, or film viewing conditions can decrease the sensitivity of mammography. To establish and maintain these optimal conditions, it is very important to perform a thorough acceptance evaluation of the equipment and the film processing conditions to determine baseline values and to monitor performance levels with periodic quality control testing.

Ultimate responsibility for mammography quality assurance rests with the radiologist in charge of the mammography practice, who must ensure that all inter-

TABLE 8-8. PERIODIC TESTS PERFORMED BY THE QUALITY CONTROL (QC) TECHNOLOGIST

Test and Frequency	Requirements for Acceptable Operation	Documentation Guidance	Timing of Corrective Action
Processor base + fog density—daily	$\leq +0.03$ OD of established operating level	QC records and charts for last 12 months; QC films for 30 days	Before any further clinical films are processed
Processor mid-density (MD) value—daily	Within ± 0.15 OD of established operating level for MD		
Processor density difference (DD) value—daily	Within ± 0.15 OD of established operating level for DD		
Phantom image center OD and reproducibility—weekly	Established operating level OD ≥ 1.20 within ± 0.20 OD of established operating level at typical clinical setting	QC records and charts for last 12 months; phantom images for the last 12 weeks	Before any further exams are performed using the x-ray unit
Phantom density difference background and test object	Within ± 0.05 OD of established operating level		
Phantom scoring—weekly	Minimum score of four fibers, three speck groups, three masses		
Fixer retention in film—quarterly	Residual fixer no greater than $5 \mu\text{g}/\text{cm}^2$	QC records since the last inspection or the past three tests	Within 30 days of the date of the test
Repeat analysis—quarterly	Operating level for repeat rate $< 2\%$ change (up or down) from previous rate		
Darkroom fog—semiannually	OD difference ≤ 0.05	QC records since last inspection or the past three tests; fog QC films from the previous three tests	Before any further clinical films are processed
Screen-film (S-F) contact—semiannually	All mammography cassettes tested with a 40-mesh copper screen, with no obvious artifacts	QC records since last inspection or the past three tests; S-F contact QC films from the previous three tests	Before any further examinations are performed using the cassettes
Compression device—semiannually	Compression force ≥ 11 newtons (25 pounds)	QC records since the last inspection or the past three tests	Before examinations are performed using the compression device

TABLE 8-9. SUMMARY TABLE OF ANNUAL QUALITY CONTROL TESTS PERFORMED BY THE PHYSICIST

Test	Regulatory Action Levels	Required Documentation	Timing of Corrective Action
AEC	OD exceeds the mean by more than 0.30 (over 2–6 cm range) or the phantom image density at the center is less than 1.20	The two most recent survey reports	Within 30 days of the date of the test
kVp	Exceeds 5% of indicated or selected kVp; C.O.V. exceeds 0.02		
Focal spot	See Table 8.1		
HVL	See Tables 8.2 and 8.3		
Air kerma (exposure) and AEC reproducibility	Reproducibility C.O.V. ^a exceeds 0.05		Before any further examinations are performed using the x-ray machine
Dose	Exceeds 3.0 mGy (0.3 rad) per exposure		
X-ray field/light field/compression device alignment	Exceeds 2% SID at chest wall Paddle visible on image		
Screen speed uniformity	O.D. variation exceeds 0.30 from the maximum to the minimum		
System artifacts	Determined by physicist		
Radiation output	Less than 4.5 mGy/sec ^b (513 mR/sec)		
Automatic decompression control	Failure of override or manual release		Before any further examinations are performed
Any applicable annual new modality tests	To be determined		

^aC.O.V., coefficient of variation, equal to the standard deviation of a series of measurements divided by the mean value.

AEC, automatic exposure control; SID, source-to-image distance.

^bNew regulations in effect on October 28, 2002 require a minimum of 700 mR/sec.

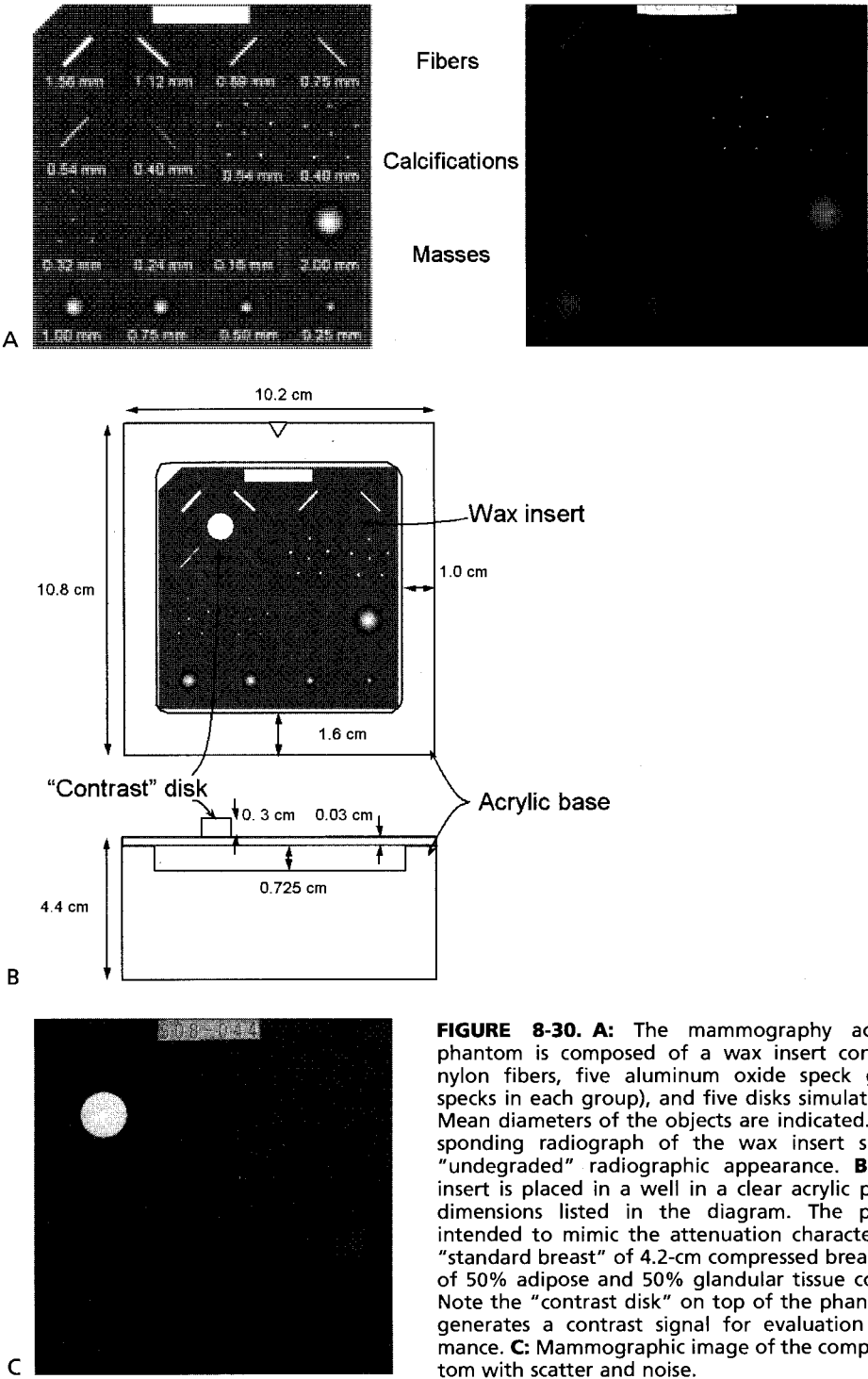


FIGURE 8-30. A: The mammography accreditation phantom is composed of a wax insert containing six nylon fibers, five aluminum oxide speck groups (six specks in each group), and five disks simulating masses. Mean diameters of the objects are indicated. The corresponding radiograph of the wax insert shows their “undegraded” radiographic appearance. **B:** The wax insert is placed in a well in a clear acrylic phantom of dimensions listed in the diagram. The phantom is intended to mimic the attenuation characteristics of a “standard breast” of 4.2-cm compressed breast thickness of 50% adipose and 50% glandular tissue composition. Note the “contrast disk” on top of the phantom, which generates a contrast signal for evaluation of performance. **C:** Mammographic image of the composite phantom with scatter and noise.

preting radiologists, mammography technologists, and the medical physicist meet the initial qualifications and maintain the continuing education and experience required by the MQSA regulations. Records must be maintained of employee qualifications, quality assurance, and the physicist's tests. Clinical performance and quality control issues should be periodically reviewed with the entire mammography staff.

The technologist performs the actual examinations, i.e., patient positioning, compression, image acquisition, and film processing. The technologist also performs the daily, weekly, monthly, and semiannual quality control tests. The medical physicist is responsible for equipment performance measurements before first clinical use, annually thereafter, and following repairs, and for oversight of the quality control program performed by the technologist.

SUGGESTED READING

- American College of Radiology. Mammography quality control manual, 1999. American College of Radiology publication PQAM99. Reston, VA: ACR, 1999. Document is available at <http://www.acr.org>.
- Barnes GT, Frey GD, eds. *Screen-film mammography: imaging considerations and medical physics responsibilities*. Madison, WI: Medical Physics, 1991.
- Food and Drug Administration Mammography Program (MQSA). The MQSA regulations and other documentation are available on the Web at <http://www.fda.gov/cdrh/mammography>.
- Haus AG, Yaffe MJ, eds. *Syllabus: a categorical course in physics: technical aspects of breast imaging*. Oak Brook, IL: RSNA, 1994.

FLUOROSCOPY

Fluoroscopy is an imaging procedure that allows real-time x-ray viewing of the patient with high temporal resolution. Before the 1950s, fluoroscopy was performed in a darkened room with the radiologist viewing the faint scintillations from a thick fluorescent screen. Modern fluoroscopic systems use image intensifiers coupled to closed-circuit television systems. Image-intensified fluoroscopy has experienced many technologic advancements in recent years. The image intensifier (II) has increased in size from the early 15-cm (6-inch) diameter systems to 40-cm (16-inch) systems available today. Modern television (TV) cameras are superior in quality to, and higher in resolution than, earlier cameras, and digital image technology is intrinsic to most modern fluoroscopy systems. New features such as variable frame rate pulsed fluoroscopy provide improved x-ray dose efficiency with better image quality. And while II-based fluoroscopy has matured technologically, flat panel detectors based on thin film transistor (TFT) technology are just over the horizon, poised to compete with and eventually replace image intensifiers.

9.1 FUNCTIONALITY

Real-Time Imaging

“Real-time” imaging is usually considered to be 30 frames per second, which is the standard television frame rate in the United States. Most general-purpose fluoroscopy systems use television technology, which provides 30 frames per second imaging. Fluoroscopic image sequences are typically not recorded, but when recording fluoroscopy is necessary (e.g., barium swallow examinations), high-quality videotape recorders can be used for playback and interpretation. Newer fluoroscopy systems allow the acquisition of a real-time digital sequence of images (digital video), that can be played back as a movie loop. Unrecorded fluoroscopy sequences are used for advancing catheters during angiographic procedures (*positioning*), and once the catheter is in place, a sequence of images is recorded using high frame rate pulsed radiography as radiographic contrast media are injected in the vessels or body cavity of interest. In gastrointestinal fluoroscopy, much of the diagnosis is formulated during the unrecorded fluoroscopy sequence; however, radiographic images are acquired for documenting the study, and to show important diagnostic features. For recording images of the heart, cine cameras offer up to 120-frame-per-second acquisition rates using 35-mm cine film. Digital cine is also available and gaining favor in the cardiology community. Additional imaging devices associated

with the fluoroscopy imaging chain such as film photo-spot, digital photo-spot, or spot film devices provide the hard-copy acquisition capability for documenting the entire range of fluoroscopic procedures.

9.2 FLUOROSCOPIC IMAGING CHAIN COMPONENTS

A standard fluoroscopic imaging chain is shown in Fig. 9-1. The x-ray tube, filters, and collimation are similar technologies to those used in radiography and are not discussed in detail here. The principal component of the imaging chain that distinguishes fluoroscopy from radiography is the image intensifier. The image output of a fluoroscopic imaging system is a projection radiographic image, but in a typical 10-minute fluoroscopic procedure a total of 18,000 individual images are produced. Due to the sheer number of images that must be produced to depict motion, for radiation dose reasons, fluoroscopic systems should produce a usable image with relatively few x-ray photons. Consequently, a very sensitive detector is needed. Image intensifiers are several thousand times more sensitive than a standard 400-speed screen-film cassette, and in principle can produce images using several thousand times less radiation. In practice, standard fluoroscopy uses about 1 to 5 μR incident upon the image intensifier per image, whereas a 400-speed screen-film system requires an exposure of about 600 μR to achieve an optical density of 1.0.

The Image Intensifier

A diagram of a modern II is shown in Fig. 9-2. There are four principal components of an II: (a) a vacuum bottle to keep the air out, (b) an input layer that converts the

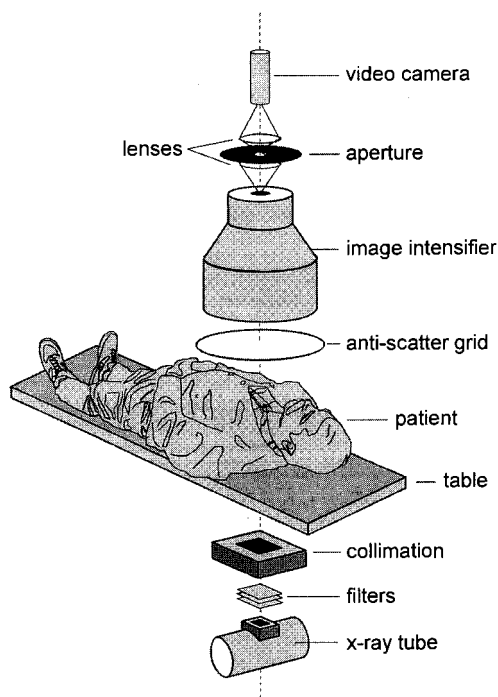


FIGURE 9-1. The fluoroscopic imaging chain with key components indicated.

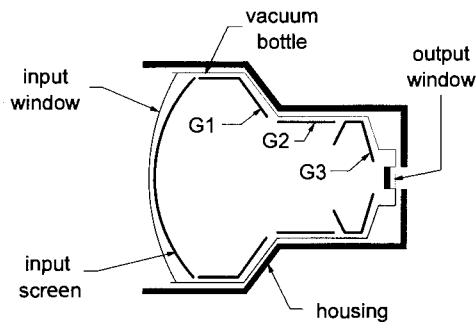


FIGURE 9-2. The internal structure of an image intensifier. The photocathode (in the input screen), the three focusing electrodes (G1, G2, and G3), and the anode (part of the output window) make up the five-element (pentode) electron optical system of the II.

x-ray signal to electrons, (c) electronic lenses that focus the electrons, and (d) an output phosphor that converts the accelerated electrons into visible light.

Input Screen

The input screen of the II consists of four different layers, as shown in Fig. 9-3. The first layer is the vacuum window, a thin (typically 1 mm) aluminum (Al) window that is part of the vacuum bottle. The vacuum window keeps the air out of the II, and its curvature is designed to withstand the force of the air pressing against it. The vacuum window of a large field-of-view (FOV) (35-cm) II supports over a ton of air pressure. A vacuum is necessary in all devices in which electrons are accelerated across open space. The second layer is the support layer, which is strong enough to support the input phosphor and photocathode layers, but thin enough to allow most x-rays to pass through it. The support, commonly 0.5 mm of aluminum, is the first component in the electronic lens system, and its curvature is designed for accurate electronic focusing.

After passing through the Al input window and the Al substrate, x-rays strike the input phosphor, whose function is to absorb the x-rays and convert their energy into visible light. The input phosphor in an II serves a purpose similar to that of the intensifying screen in screen-film radiography, and is subject to the same compromise—it must be thick enough to absorb the majority of the incident x-rays but

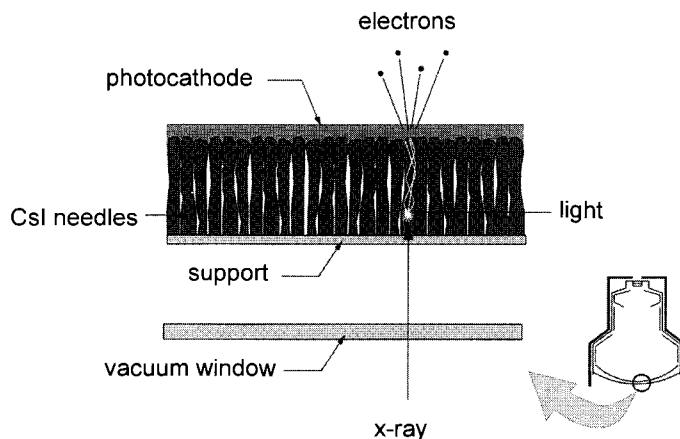


FIGURE 9-3. The input section of an image intensifier (II). X-rays must pass through the vacuum window and support, before striking the cesium iodide (CsI) input phosphor. CsI forms in long crystalline needles that act like light pipes, limiting the lateral spread of light and preserving spatial resolution. Light emitted in the cesium iodide strikes the photocathode, causing electrons to be liberated into the electronic lens system of the II.

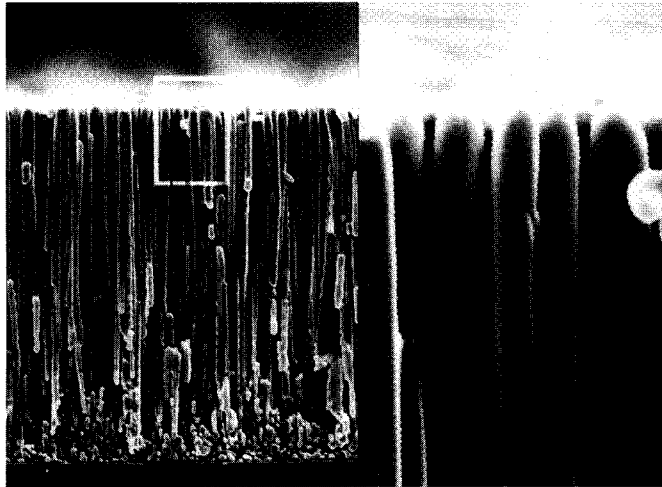


FIGURE 9-4. A scanning electron micrograph illustrates the needle-like structure of a CsI input phosphor.

thin enough not to degrade the spatial resolution of the image significantly. IIs offer a high vacuum environment that a screen-film cassette cannot, and therefore unique phosphor technology can be used. Virtually all modern IIs use cesium iodide (CsI) for the input phosphor. CsI is not commonly used in screen-film cassettes because it is hygroscopic and would degrade if exposed to air. CsI has the unique property of forming in long, needle-like crystals. The long, thin columns of CsI function as light pipes, channeling the visible light emitted within them toward the photocathode with minimal lateral spreading. As a result, the CsI input phosphor can be quite thick and still maintain high resolution. The CsI crystals are approximately 400 μm tall and 5 μm in diameter, and are formed by vacuum evaporation of CsI onto the substrate (Fig. 9-4). For a 60-keV x-ray photon absorbed in the phosphor, approximately 3,000 light photons (at about 420 nm wavelength) are emitted. The *K*-edges of cesium (36 keV) and iodine (33 keV) are well positioned with respect to the fluoroscopic x-ray spectrum, and this fact helps contribute to a high x-ray absorption efficiency.

The photocathode is a thin layer of antimony and alkali metals (such as Sb_2S_3) that emits electrons when struck by visible light. With a 10% to 20% conversion efficiency, approximately 400 electrons are released from the photocathode for each 60-keV x-ray photon absorbed in the phosphor.

Electron Optics

Once x-rays are converted to light and then to electrons in the input screen, the electrons are accelerated by an electric field. The energy of each electron is substantially increased, and this gives rise to *electronic gain*. The spatial pattern of electrons released at the input screen must be maintained at the output phosphor (although minified), and therefore electron focusing is required. Focusing is achieved using an electronic lens, which requires the input screen to be a curved surface, and this results in unavoidable *pincushion distortion* of the image (Fig. 9-5). The G1, G2, and G3 electrodes illustrated in Fig. 9-2, along with the input screen substrate (the cathode) and the anode near the output phosphor comprise the five-component

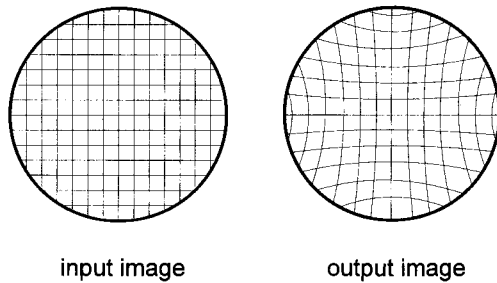


FIGURE 9-5. Because the input surface of the image intensifier is curved, and the output surface is flat, *pincushion distortion* results. For a true rectilinear grid input to the system (**left**), the output image will demonstrate spatial distortion (**right**).

(“pentode”) electronic lens system of the II. The electrons are released from the photocathode with very little kinetic energy, but under the influence of the ~25,000 to 35,000 V electric field, they are accelerated and arrive at the anode with high velocity and considerable kinetic energy. The intermediate electrodes (G1, G2, and G3) shape the electric field, focusing the electrons properly onto the output layer. After penetrating the very thin anode, the energetic electrons strike the output phosphor and cause a burst of light to be emitted.

The Output Phosphor

The output phosphor (Fig. 9-6) is made of zinc cadmium sulfide doped with silver (ZnCdS:Ag), which has a green (~530 nm) emission spectrum. Green light is in the middle of the visible spectrum and is well matched to the spectral sensitivity of many video camera target materials. The ZnCdS phosphor particles are very small (1 to 2 μm), and the output phosphor is quite thin (4 to 8 μm), to preserve high spatial resolution. The anode is a very thin (0.2 μm) coating of aluminum on the vacuum side of the output phosphor, which is electrically conductive to carry away the electrons once they deposit their energy in the phosphor. Each electron causes the emission of approximately 1,000 light photons from the output phosphor.

The image is much smaller at the output phosphor than it is at the input phosphor, because the 23- to 35-cm diameter input image is focused onto a circle with a 2.5-cm diameter. To preserve a resolution of 5 line pairs/mm at the input plane, the output phosphor must deliver resolution in excess of 70 line pairs/mm. This is why a very fine grain phosphor is required at the output phosphor. The reduction in image diameter also leads to amplification, since the energy incident on the 415

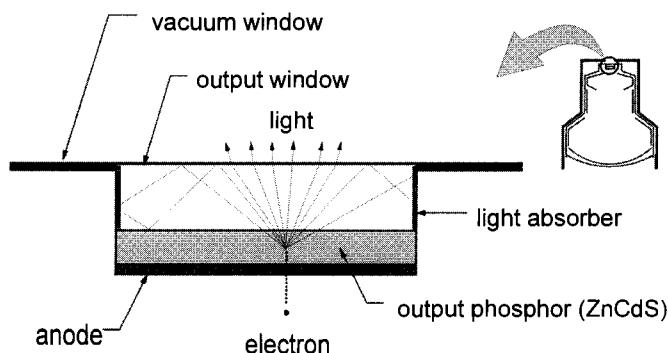


FIGURE 9-6. The output window of an image intensifier. The electrons strike the output phosphor, causing the emission of light. The thick glass output window is part of the vacuum housing and allows light to escape the top of the II. Light that is reflected in the output window is scavenged to reduce glare by the addition of a light absorber around the circumference of the output window.

cm^2 area of the 9-inch-diameter input phosphor is concentrated to the 5 cm^2 output phosphor. By analogy, a magnifying glass collects sunlight over its lens surface area and focuses it onto a small spot, and in the process the light intensity is amplified enough to start a fire. The so-called *minification gain* of an image intensifier is simply the ratio of the area of the input phosphor to that of the output phosphor. The ratio of areas is equal to the square of the diameter ratio, so a 9-inch II (with a 1-inch output phosphor) has a minification gain of 81, and in 7-inch mode it is 49.

The last stage in the II that the image signal passes through is the output window (Fig. 9-6). The output window is part of the vacuum bottle, and must be transparent to the emission of light from the output phosphor. The output phosphor is coated right onto the output window. Some fraction of the light emitted by the output phosphor is reflected at the glass window. Light bouncing around the output window is called *veiling glare*, and can reduce image contrast. This glare is reduced by using a thick (about 14 mm) clear glass window, in which internally reflected light eventually strikes the side of the window, which is coated black to absorb the scattered light.

Characteristics of Image Intensifier Performance

The function of the x-ray II is to convert an x-ray image into a minified light image. There are several characteristics that describe how well the II performs this function. These parameters are useful in specifying the capabilities of the II, and are also useful in troubleshooting IIs when they are not performing properly.

Conversion Factor

The conversion factor is a measure of the gain of an II. The input to the II is an x-ray exposure rate, measured in mR/sec . The output of the II is luminance, measured in candela per meter squared. The conversion factor of an II is the ratio of the (output) luminance divided by the (input) exposure rate, resulting in the peculiar units of $\text{Cd sec m}^{-2} \text{ mR}^{-1}$. The conversion factor of a new II ranges from 100 to 200 $\text{Cd sec m}^{-2} \text{ mR}^{-1}$. The conversion factor degrades with time, and this ultimately can lead to the need for II replacement.

Brightness Gain

The brightness gain is the product of the electronic and minification gains of the II. The electronic gain of an II is roughly about 50, and the minification gain changes depending on the size of the input phosphor and the magnification mode. For a 30-cm (12-inch) II, the minification gain is 144, but in 23-cm (9-inch) mode it is 81, and in 17-cm (7-inch) mode it is 49. The brightness gain therefore ranges from about 2,500 to 7,000. As the effective diameter of the input phosphor decreases (increasing magnification), the brightness gain *decreases*.

Field of View/Magnification Modes

Image intensifiers come in different sizes, commonly 23-, 30-, 35-, and 40-cm (9-, 12-, 14-, and 16-inch) fields of view (FOV). Large IIs are useful for gastrointestinal/genitourinary (GI/GU) work, where it is useful to cover the entire abdomen.

For cardiac imaging, by comparison, the 23-cm (9-inch) image intensifier is adequate for imaging the heart, and its smaller size allows tighter positioning. In addition to the largest FOV, which is determined by the physical size of the II, most IIs have several *magnification modes*. Magnification is produced (Fig. 9-7) by pushing a button that changes the voltages applied to the electrodes in the II, and this results in different electron focusing. As the magnification factor increases, a smaller area on the input of the II is visualized. When the magnification mode is engaged, the collimator also adjusts to narrow the x-ray beam to the smaller field of view.

As discussed above, the brightness gain of the image intensifier decreases as the magnification increases. The automatic brightness control circuitry (discussed later) compensates for the dimmer image by boosting the x-ray exposure rate. The increase in the exposure rate is equal to the ratio of FOV areas. Take for example a 30-cm (12-inch) II, which has 23-cm (9-inch) and 18-cm (7-inch) magnification modes. Switching from the 12-inch mode to the 9-inch mode will increase the x-ray exposure rate by a factor of $(12/9)^2 = 1.8$, and going from the 12-inch to the 7-inch mode will increase the exposure rate by a factor of $(12/7)^2 = 2.9$. Therefore, as a matter of radiation safety, the fluoroscopist should use the largest field of view (the least magnification) that will facilitate the task at hand.

Contrast Ratio

The *contrast ratio* is an indirect measurement of the veiling glare of an image intensifier. With a constant input x-ray exposure rate to the II, the light intensity at the center of the output phosphor is measured using a light meter. A thick, 2.5-cm-diameter lead disk is then placed in the center of the input phosphor, blocking the radiation to the input and theoretically cutting the light at the center of the output phosphor to zero. However, light produced in the periphery of the II will scatter, and consequently some light intensity will be measured at the center of the output phosphor. The contrast ratio is simply the ratio of light intensity, with and without the lead disk being present. Contrast ratios of 15 to 30 are commonly found.

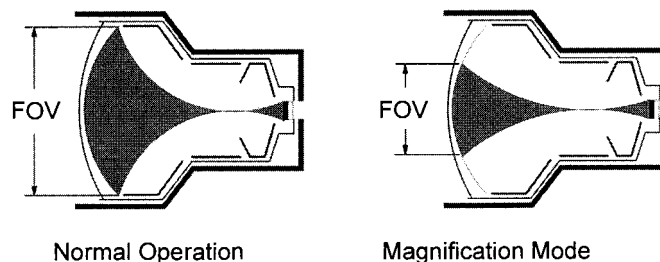


FIGURE 9-7. In normal operation of the image intensifier (left), electrons emitted by the photocathode over the entire surface of the input window are focused onto the output phosphor, resulting in the maximum field of view (FOV) of the II. Magnification mode (right) is achieved by pressing a button that modulates the voltages applied to the five electrodes, which in turn changes the electronic focusing such that only electrons released from the smaller diameter FOV are properly focused onto the output phosphor.

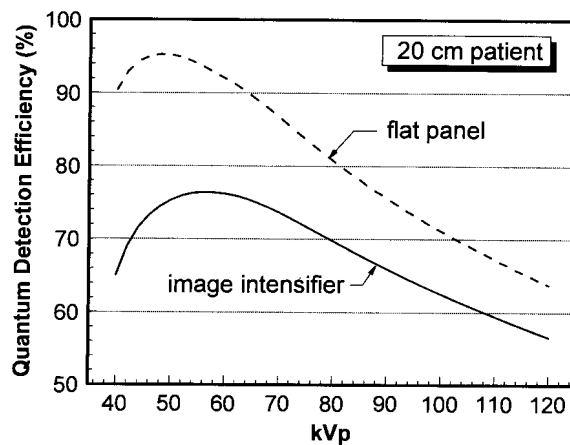


FIGURE 9-8. The quantum detection efficiency (QDE) of an image intensifier as a function of kVp. The x-ray beam was hardened by a 20-cm-thick patient for this calculation. The QDE for the image intensifier is reduced by the ~1.5 mm of aluminum support structures in front of the input phosphor. Flat panel imaging systems based on thin film transistor technology require only a carbon-fiber face plate for protection and light-tightness, and therefore demonstrate slightly higher QDEs.

Quantum Detection Efficiency

X-rays must pass through the vacuum window (~1.0 mm Al) and the input screen substrate (~0.5 mm Al) before reaching the CsI input phosphor (~180 mg/cm²). Figure 9-8 illustrates the kVp-dependent quantum detection efficiency (QDE) of an II. The quantum detection efficiency is maximal around 60 kVp; however, the dose to the patient will decrease at higher kVps, so the optimal kVp for a given examination will generally be higher than 60 kVp.

S Distortion

S distortion is a spatial warping of the image in an S shape through the image. This type of distortion is usually subtle, if present, and is the result of stray magnetic fields and the earth's magnetic field. On fluoroscopic systems capable of rotation, the position of the S distortion can shift in the image due to the change in the system's orientation with respect to the earth's magnetic field.

Optical Coupling

Distribution Mechanism

The output image on an II is small, and for over-table II's, a ladder would be needed to view the output window directly. Consequently, a video camera is usually mounted above the II and is used to relay the output image to a TV monitor for more convenient viewing by the operator. In addition to the TV camera, other image recording systems are often connected to the output of the II. The *optical distributor* (Fig. 9-9) is used to couple these devices to the output image of the II. A lens is positioned at the output image of the II, and light rays travel in a parallel (nondiverging) beam into the light-tight distributor housing. A mirror or a prism is used to reflect or refract the light toward the desired imaging receptor. Figure 9-9 illustrates a video camera and a 105-mm roll film camera mounted on the distributor. Mirrors can be motorized, and can shunt the light to any of the accessory ports on the distributor. When only two imaging receptors are present, a stationary

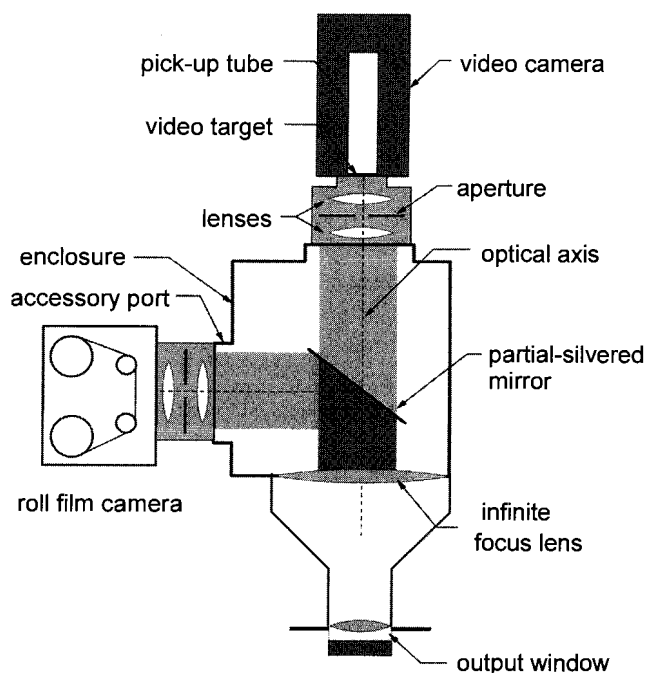


FIGURE 9-9. The optical distributor. The output window of the image intensifier, which is the source of the optical image, is also shown (*bottom*). Parallel rays of light enter the optical chamber, are focused by lenses, and strike the video camera where an electronic image is produced. A partially silvered mirror (or motorized mirror in some instances) is used to shunt the light emitted by the image intensifier to an accessory port (*left*). A roll film camera is shown mounted on the optical distributor. The separate apertures for the two imaging devices (video camera and roll film camera) allow adjustment for the differing sensitivity of the two devices.

partially silvered mirror can be used to split the light beam so that it is shared between the two cameras.

Lenses

The lenses used in fluoroscopy systems are identical to high-quality lenses used in photography. The lenses focus the incoming light onto the focal plane of the camera. The lens assemblies include a variable aperture, basically a small hole positioned between individual lenses in the lens assembly. Adjusting the size of the hole changes how much light gets through the lens system. The *f-number* is (inversely) related to the *diameter* of the hole ($f = \text{focal length/aperture diameter}$), but it is the *area* of the hole that determines how much light gets through. Because of this, the standard *f-numbers* familiar to anyone who owns a camera progress by multiples of $\sqrt{2}$: 1.0, 1.4, 2.0, 2.8, 4.0, 5.6, 8, 11, 16. Changing the diameter of the hole by a factor $\sqrt{2}$ changes its area by a factor of 2, and thus increasing the *f-number* by one *f-stop* reduces the amount of light passing through the aperture by a factor of 2. At $f/16$, the aperture is tiny, little light gets through, and the overall gain of the II and optics combined is reduced. At $f/5.6$, more light gets through, and the overall gain of the II-optics subsystem is eight times higher than at $f/16$.

Adjustment of the aperture markedly changes the effective gain of the II-optics subcomponents of the imaging chain. This adjustment has an important affect on the performance of the fluoroscopy system. By lowering the gain of the II optics, a higher x-ray exposure rate is used, and lower noise images will result. Increasing the gain reduces the x-ray exposure rate and lowers the dose, but reduces image quality. This will be discussed more fully below (see Automatic Brightness Control).

Video Cameras

General Operation

The closed-circuit video system transfers the signal from the output of the II to the video monitors in the room. The operation of a video system is shown schematically in Fig. 9-10. In the video camera, patterns of light (the image data) are incident upon the TV target (Fig. 9-10A). The target is swept by a scanning electron beam in a raster scan pattern. The TV target is made of a photoconductor, which has high electrical resistance in the dark, but becomes less resistive as the light intensity striking it increases. As the electron beam (electrical current) scans each location on the target, the amount of current that crosses the TV target and reaches the signal plate depends on the resistance of the target at each location, which in turn is related to the local light intensity. Thus the electrical signal is modulated by the

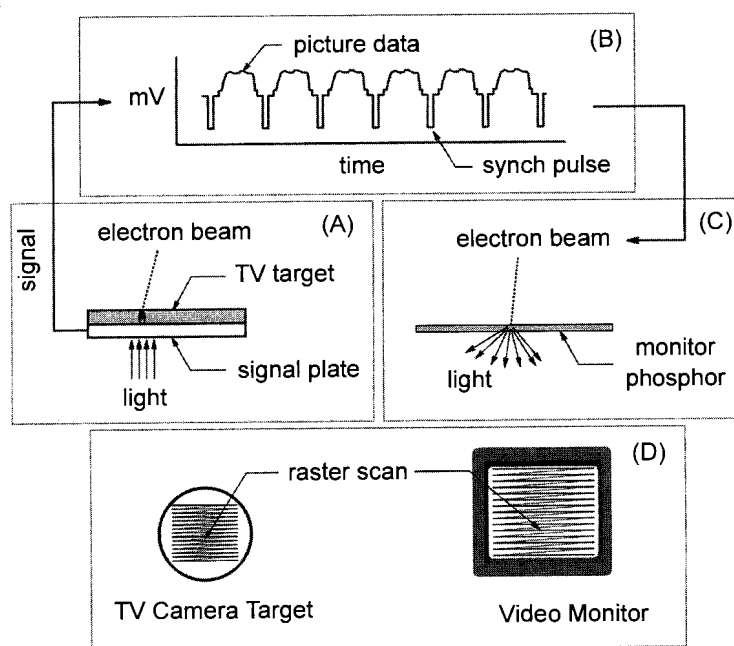


FIGURE 9-10. The closed circuit TV system used in fluoroscopy. At the TV camera (A), an electron beam is swept in raster fashion on the TV target (e.g., SbSO_3). The TV target is a photoconductor, whose electrical resistance is modulated by varying levels of light intensity. In areas of more light, more of the electrons in the electron beam pass across the TV target and reach the signal plate, producing a higher video signal in those lighter regions. The video signal (B) is a voltage versus time waveform that is communicated electronically by the cable connecting the video camera with the video monitor. Synchronization pulses are used to synchronize the raster scan pattern between the TV camera target and the video monitor. Horizontal sync pulses (shown) cause the electron beam in the monitor to laterally retrace and prepare for the next scan line. A vertical sync pulse (not shown) has different electrical characteristics and causes the electron beam in the video monitor to reset at the top of the screen. Inside the video monitor (C), the electron beam is scanned in raster fashion, and the beam current is modulated by the video signal. Higher beam current at a given location results in more light produced at that location by the monitor phosphor. (D) The raster scan on the monitor is done in synchrony with the scan of the TV target.

local variations in light intensity, and that is physically how the video system converts the image to an electronic signal.

The video signal, which is represented as voltage versus time (Fig. 9-10B), is transferred by wire to the video monitor. Synchronization pulses are electronically added to the signal to keep the raster scanning in the video camera and video monitor synchronized. At the monitor, the raster scan pattern on the target is replicated by an electron beam scanning the monitor phosphor. The video signal voltage modulates the beam current in the monitor, which in turn modulates the luminance at that location on the monitor.

Analog video systems typically have 30 frame/sec operation, but they work in an *interlaced* fashion to reduce *flicker*, the perception of the image flashing on and off. The human eye-brain system can detect temporal fluctuations slower than about 47 images/sec, and therefore at 30 frame/sec flicker would be perceptible. With interlaced systems, each frame is composed of two fields (called *odd* and *even* fields, corresponding to every other row in the raster, with the odd field starting at row 1, and the even field starting at row 2), and each field is refreshed at a rate of 60 times per second (although with only half the information), which is fast enough to avoid the perception of flicker.

Video Resolution

The spatial resolution of video in the vertical direction is determined by the number of video lines. Standard video systems use 525 lines (in the United States), but larger IIs (>23 cm) require *high line rate* video to deliver adequate resolution, which is typically 1,023 lines, but this varies by vendor. For a 525-line system, only about 490 lines are usable. In the early days of television, a man named Kell determined empirically that about 70% of theoretical video resolution is appreciated visually, and this psychophysical effect is now called the Kell factor. Thus, of the 490 video lines, about $490 \times 0.7 = 343$ are useful for resolution, and 343 *lines* are equivalent to about 172 *line pairs*, since one line pair constitutes two lines. If 172 line pairs scan a 229-mm (9-inch) field, the resulting spatial resolution will be 172 line pairs/229 mm = 0.75 line pairs/mm. In 17-cm (7-inch) and 12-cm (5-inch) modes, this calculation yields resolutions of 1.0 and 1.4 line pairs/mm, respectively.

The horizontal resolution is determined by how fast the video electronics can respond to changes in light intensity. This is influenced by the camera, the cable, and the monitor, but the horizontal resolution is governed by the *bandwidth* of the system. The time necessary to scan each video line (525 lines at 30 frames/sec) is 63 μ sec. For standard video systems, 11 μ sec are required for the horizontal retrace (the repositioning of the electron beam at the beginning of each horizontal raster scan), so that 52 μ sec are available for portraying the picture data. To achieve 172 cycles in 52 μ sec (to match the vertical resolution), the bandwidth required is 172 cycles / 52×10^{-6} sec = 3.3×10^6 cycles/sec = 3.3 MHz. Higher bandwidths ($\sim 4\times$) are required for high-line rate video systems.

Flat Panel Digital Fluoroscopy

Flat panel devices are thin film transistor (TFT) arrays that are rectangular in format and are used as x-ray detectors. TFT systems are pixelated devices with a

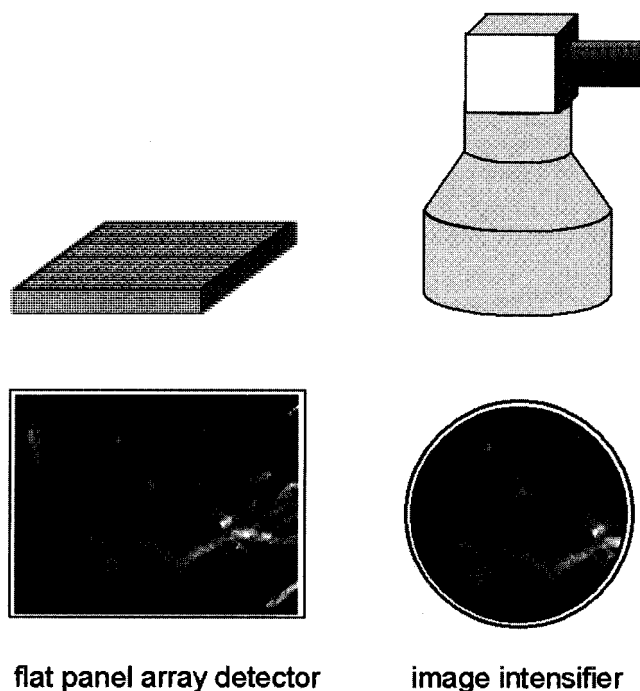


FIGURE 9-11. The flat panel imaging system and the image intensifier with its mounted video camera and optimal distributor. The flat panel detector produces a rectangular image, well matched to the rectangular format of TV monitors. The image produced by the image intensifier is circular in format, resulting in less efficient utilization of rectangular monitors for fluoroscopic display. The flat panel detector is substantially less bulky than the image intensifier and TV system, but provides the same functionality.

photodiode at each detector element, which converts light energy to an electronic signal. Since the TFT array is sensitive to visible light (and is not very sensitive to x-rays), a scintillator such as CsI is used to convert the incident x-ray beam energy into light, which then strikes the TFT detector. A more complete discussion of TFT arrays is given in Chapter 11. For fluoroscopic applications, the flat panel imager replaces the image intensifier and video camera, and directly records the real-time fluoroscopic image sequence. The pixel size in fluoroscopy is usually larger than radiography, and some flat panel systems have the ability to adjust the pixel size by binning four pixels into one larger pixel. Such dual-use systems have pixels small enough to be adequate for radiography (e.g., 100 to 150 μm), but the pixels can be binned to provide a detector useful for fluoroscopy (e.g., 200 to 300 μm). Flat panel imagers (Fig. 9-11) can therefore replace the II, video camera, digital spot film device, and cine camera in a much lighter and smaller package. Because a vacuum environment is not required (because there are no electron optics), the cover to the flat panel can be ~ 1 mm of carbon fiber, and this improves the quantum detection efficiency compared to the image intensifier as shown in Fig. 9-8.

9.3 PERIPHERAL EQUIPMENT

There are many imaging cameras that are commonly used with fluoroscopic systems. Most of these are mounted on accessory ports of the optical distributor (Fig. 9-9).

Photo-Spot Cameras

A photo-spot camera is used to generate images on photographic film, taken from the output of the II. The most common formats are 100-mm cut film, or 105-mm roll film. Spot film cameras are convenient; fluoroscopists can archive an image during the fluoroscopic examination by simply stepping on a floor pedal. The mirror in the optical distributor swings into the optical path, redirecting light rays toward the spot-film camera (or a partially silvered mirror is already in place). The x-ray generator produces a radiographic pulse of x-rays, the camera shutter opens briefly, and the image is recorded on film. The film is advanced and stored in a take-up magazine. Photo-spot images are made directly from the output phosphor of the II, through lenses, and therefore enjoy the full resolution of the II systems, unencumbered by the resolution constraints of the video system. An entrance exposure to the II of 75 to 100 $\mu\text{R}/\text{image}$ is typically required for photo-spot cameras operating with a 23-cm (9-inch) FOV II.

Digital Photo-Spot

Digital photo-spot cameras are high-resolution, slow-scan TV cameras in which the TV signal is digitized and stored in computer memory. Alternately, digital photo-spot cameras are charge coupled device (CCD) cameras with 1024^2 or 2048^2 pixel formats. The digital photo-spot camera allows near-instantaneous viewing of the image on a video monitor. The real-time acquisition of the permanent archive images allows the fluoroscopist to quickly assemble an array of images demonstrating the anatomy pertinent to the diagnosis. The digital images can be printed on a laser imager if hard-copy viewing is necessary.

Spot-Film Devices

A spot-film device attaches to the front of the II, and allows the acquisition of radiographic screen-film images using a few button presses on the device. When an image acquisition is requested by the operator, the spot-film device transports a screen-film cassette from a lead-lined magazine to a position directly in front of the II. A radiographic pulse of x-rays is used, and spot-film devices typically have their own antiscatter grids and automatic exposure control. Most systems automatically collimate to a variety of formats. For example, one 30-cm $\times$ 30-cm cassette can be exposed with four different 15-cm $\times$ 15-cm radiographic images. Although bulky, spot-film devices are convenient when large FOV images are routinely required, as in GI/GU studies. Spot film devices produce conventional screen-film images, with somewhat better spatial resolution than images produced by the II (e.g., photo-spot cameras).

Cine-Radiography Cameras

A cine camera attaches to a port on the optical distributor of an II, and can record a very rapid sequence of images on 35-mm film. Cine is used frequently in cardiac studies, where a very high frame rate is needed to capture the rapid motion of the heart. The frame rates of cine cameras run from 30 frames/sec to 120 frames/sec or

higher. Cine radiography uses very short radiographic pulses, and therefore special generators are needed. The opening of the cine camera shutter is synchronized with the x-ray pulses. The x-ray tube loading is usually quite high, so high heat capacity x-ray tubes are used in the cardiac catheterization laboratory. An entrance exposure to the input phosphor of approximately 10 to 15 $\mu\text{R}/\text{frame}$ (23-cm-diameter II) is used for cine-radiography studies. Digital cine cameras are typically CCD-based cameras that produce a rapid sequence of digital images instead of film sequence. Digital cine is increasingly used in the cardiology community.

9.4 FLUOROSCOPY MODES OF OPERATION

With the computerization of x-ray generators and imaging systems, a great deal of flexibility has been added to fluoroscopy systems compared to what was available in the 1970s. Some systems come with enhanced features, and some of these alternate modes of operation are described below.

Continuous Fluoroscopy

Continuous fluoroscopy is the basic form of fluoroscopy. It uses a continuously on x-ray beam using typically between 0.5 and 4 mA (depending on patient thickness). A video camera displays the image at 30 frames/sec, so that each fluoroscopic frame requires 33 milliseconds (1/30 second). Any motion that occurs within the 33-msec acquisition acts to blur the fluoroscopic image, but this is acceptable for most examinations. In the United States, the maximum entrance exposure to the patient is 10 R/min.

High Dose Rate Fluoroscopy

Some fluoroscopy systems have high dose rate options, usually called *specially activated fluoroscopy*. This mode of operation allows exposure rates of up to 20 R/min to the patient in the U.S. The high dose rate mode is enabled by stepping on a different pedal than the standard fluoroscopy pedal. An audible signal is required to sound when specially activated fluoroscopy is being used. Unless a practice sees an exceptionally high proportion of obese patients, the need for high dose rate fluoroscopy is questionable.

Variable Frame Rate Pulsed Fluoroscopy

In pulsed fluoroscopy, the x-ray generator produces a series of short x-ray pulses. Let's compare this with continuous fluoroscopy with 33-msec frame times, operating continuously at 2 mA. The pulsed fluoroscopy system can generate 30 pulses per second, but each pulse can be kept to ~ 10 msec and 6.6 mA. This would provide the same x-ray exposure rate to the patient (using 0.066 mAs per frame in this example), but with pulsed fluoroscopy the exposure time is shorter (10 msec instead of 33 msec), and this will reduce blurring from patient motion in the image. Therefore, in fluoroscopic procedures where object motion is high (e.g., positioning catheters in highly pulsatile vessels), pulsed fluoroscopy offers better image quality at the same dose rates.

Many fluoroscopists practice aggressive dose-saving measures, a commendable goal. Variable frame rate pulsed fluoroscopy can be instrumental in this. During much of a fluoroscopic procedure, a rate of 30 frames per second is not required to do the job. For example, in a carotid angiography procedure, initially guiding the catheter up from the femoral artery to the aortic arch does not require high temporal resolution, and perhaps 7.5 frames per second can be used—reducing the radiation dose for that portion of the study to 25% (7.5/30). Pulsed fluoroscopy at variable frame rates (typically 30, 15, and 7.5 frames/sec) allows the fluoroscopist to reduce temporal resolution when it is not needed, sparing dose in return. At low frame rates, the fluoroscopic image would demonstrate intolerable flicker unless measures were taken to compensate for this. Using a digital refresh memory, each image is digitized and shown continuously at 30 frames per second (or faster) on the video monitors until it is refreshed with the next pulsed image. In this way, a sometimes jerky temporal progression of images can be seen by the human observer at low frame rates, but the image does not suffer from on/off flicker.

Frame Averaging

Fluoroscopy systems provide excellent temporal resolution, and it is this feature that is the basis of their clinical utility. However, fluoroscopy images are also relatively noisy, and under some circumstances it is beneficial to compromise temporal resolution for lower noise images. This can be done by averaging a series of images, as shown in Fig. 9-12. Frame averaging is performed by digitizing the fluoroscopic images and performing real-time averaging in computer memory for display. Appreciable frame averaging can cause very noticeable image *lag* (reduced temporal resolution), but the noise in the image is reduced as well. The compromise between lag and image noise depends on the specific fluoroscopic application and the preferences of the user. Aggressive use of frame averaging can allow lower dose imaging in many circumstances. Portable fluoroscopy systems use x-ray tubes with limited out-

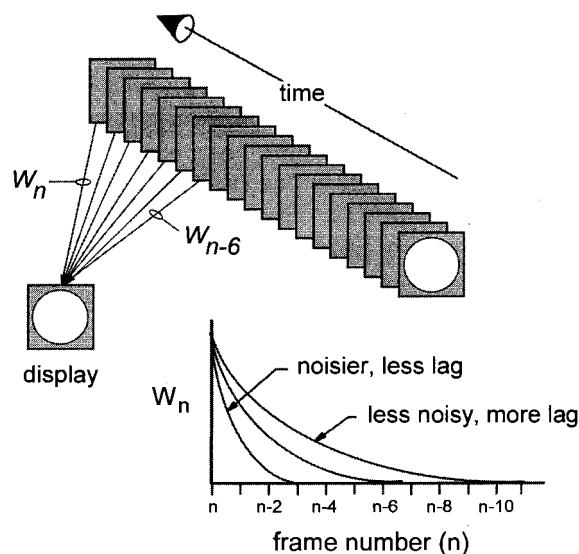


FIGURE 9-12. The concept of frame averaging. The individual frames occurring as a function of time are weighted and averaged into a single display image, and typically "older" frames are weighted less than newer frames.

put, and consequently frame averaging is a feature commonly found on these systems.

Different averaging algorithms can be used, but a common approach to frame averaging is called *recursive filtering*. In this approach, the image just acquired, I_n , is added with the last image (I_{n-1}) using

$$I_{displayed} = \alpha I_n + (1 - \alpha) I_{(n-1)} \quad [9-1]$$

The parameter α ranges from 0 to 1; as α is reduced, the contribution of the current image (I_n) is reduced, the contribution of previous images is enhanced, and thus the amount of lag is increased. As α is increased, the amount of lag is reduced, and at $\alpha = 1$, no lag occurs. The current displayed image $I_{displayed}$ becomes the I_{n-1} image for the next frame, and because of this the weighted image data from an entire sequence of images (not just two images) is included (Fig. 9-12).

Last-Frame-Hold

When the fluoroscopist takes his or her foot off of the fluoroscopy pedal, rather than seeing a blank monitor, last-frame-hold enables the last live image to be shown continuously. The last-frame-hold circuit merely is a video digitizer, a fast analog to digital converter that converts the video signal to a digital image. As the fluoroscopy pedal is released, a signal is generated that causes the last image with x-rays on to be captured, and this image is displayed continuously on the video monitor(s) until the fluoroscopy pedal is depressed again. Last-frame-hold is very convenient and can reduce the dose to the patient. While especially useful at training institutions where residents are developing their fluoroscopy skills, last-frame-hold is a convenient feature that even seasoned fluoroscopists should and do use. Rather than orienting oneself with the pedal depressed and x-rays on, last-frame-hold allows the fluoroscopist to examine the image for as long as necessary using no additional radiation.

Road Mapping

Road mapping is really a software-enhanced variant of the last-frame-hold feature. Two different approaches to road mapping can be found on fluoroscopic systems. Some systems employ side-by-side video monitors, and allow the fluoroscopist to capture an image (usually with a little contrast media injection), which is then shown on the monitor next to the live fluoroscopy monitor. In this way, the path of the vessel can be seen on one monitor, while the angiographer advances the catheter in real time on the other monitor. Another approach to road mapping is to allow the angiographer to capture a contrast injection image (or subtracted image), but then that image is displayed as an overlay onto the live fluoroscopy monitor. In this way, the angiographer has the vascular "road map" right on the fluoroscopy image, and can angle the catheter to negotiate the patient's vascular anatomy. Road mapping is useful for advancing catheters through tortuous vessels.

9.5 AUTOMATIC BRIGHTNESS CONTROL

The exposure rates in modern fluoroscopic systems are controlled automatically. The purpose of the *automatic brightness control* (ABC) is to keep the brightness of

the image constant at the monitor. It does this by regulating the x-ray exposure rate incident on the input phosphor of the II. When the II is panned from a thin to a thick region of the patient, the thicker region attenuates more of the x-rays, so fewer x-rays strike the II and less light is generated. The system senses the reduction in light output from the II and sends a signal to the x-ray generator to increase the x-ray exposure rate. Fluoroscopic systems sense the light output from the II using a *photodiode* or the video signal itself. This control signal is used in a feedback circuit to the x-ray generator to regulate the x-ray exposure rate. Automatic brightness control strives to keep the number of x-ray photons used for each fluoroscopic frame at a constant level, keeping the signal-to-noise ratio of the image approximately constant regardless of the thickness of the patient.

The fluoroscopic ABC circuitry in the x-ray generator is capable of changing the mA and the kV in continuous fluoroscopy mode. For pulsed-fluoroscopy systems, the ABC circuitry may regulate the pulse width (the time) or pulse height (the mA) of the x-ray pulses. For large patients, the x-ray exposure rate may hit the maximum legal limit (10 R/min) yet the image will still not be appropriately bright. Some x-ray fluoroscopic systems employ ABC circuitry which can then open the aperture to increase the brightness of the image. The video camera also has electronic gain, which is increased when x-ray exposure rates are at a maximum. These last two approaches will increase the visible noise in the image, since the x-ray fluence rate (photons/mm²/sec) is no longer being held constant at the input phosphor.

Any particular generator changes the mA and the kV in a predetermined sequence; however, this sequence varies with different brands of generators and is sometimes selectable on a given fluoroscopic system. How the mA and kV change as a function of patient thickness has an important effect on patient dose and image quality. When the fluoroscopist pans to a thicker region of the patient, the ABC signal requests more x-rays from the generator. When the generator responds by increasing the kV, the subject contrast decreases, but the dose to the patient is kept low because more x-rays penetrate the patient at higher kV. In situations where contrast is crucial (e.g., angiography), the generator can increase the mA instead of the kV; this preserves subject contrast at the expense of higher patient dose. In practice, the mA and kV are increased together, but the curve that describes this can be designed to aggressively preserve subject contrast, or alternately to favor a lower dose examination (Fig. 9-13). Some fluoroscopy systems have “low dose” or “high

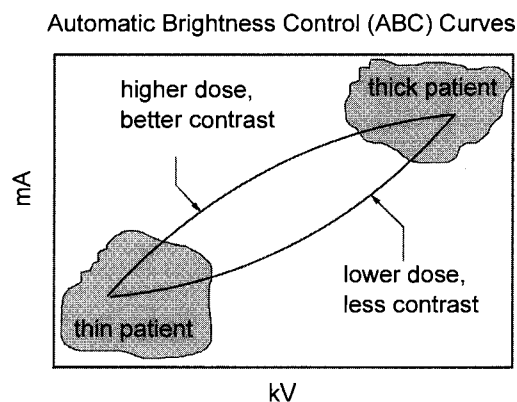


FIGURE 9-13. Automatic brightness control (ABC) changes the exposure rate as the attenuation of the patient between the x-ray tube and the image intensifier changes. The exposure rate can be increased by increasing either the mA, the kV, or both. Two possible curves are illustrated. **The top curve** increases mA more rapidly than kV as a function of patient thickness, and preserves subject contrast at the expense of higher dose. **The bottom curve** increases kV more rapidly than mA with increasing patient thickness, and results in lower dose, but lower contrast as well.

contrast" selections on the console, which select the different mA/kV curves shown in Fig. 9-13.

9.6 IMAGE QUALITY

Spatial Resolution

The *spatial resolution* of the II is best described by the modulation transfer function (MTF). The *limiting resolution* of an imaging system is where the MTF approaches zero (Fig. 9-14). The limiting resolution of modern IIs ranges between 4 and 5 cycles/mm in 23-cm mode. Higher magnification modes (smaller fields of view) are capable of better resolution; for example, a 14-cm mode may have a limiting resolution up to 7 cycles/mm. The video imaging system degrades the MTF substantially, so that the resolution performance of the image intensifier is not realized. Table 9-1 lists the typical limiting resolution of video-based fluoroscopy for various image intensifier field sizes.

Film-based imaging systems (photo-spot and cine cameras) coupled directly to the output of the image intensifier will produce spatial resolution similar to that of the image intensifier, since the spatial resolution of the optics and of the film are typically much better than that of the II. Digital imaging systems such as digital photo-spot and digital subtraction angiography (DSA) systems will usually demonstrate limiting resolution determined by the pixel size for 512^2 and 1024^2 matrix systems. For example, for a 1024^2 DSA system in 27-cm FOV with the width of the digital matrix spanning the circular FOV of the II, the pixel size would measure about $270 \text{ mm}/1024 \text{ pixels} = 0.26 \text{ mm/pixel}$. The corresponding limiting spatial resolution (Nyquist equation, see Chapter 10) is about 1.9 cycles/mm.

Contrast Resolution

The contrast resolution of fluoroscopy is low by comparison to radiography, because the low exposure levels produce images with relatively low signal-to-noise ratio (SNR). Contrast resolution is usually measured subjectively by viewing con-

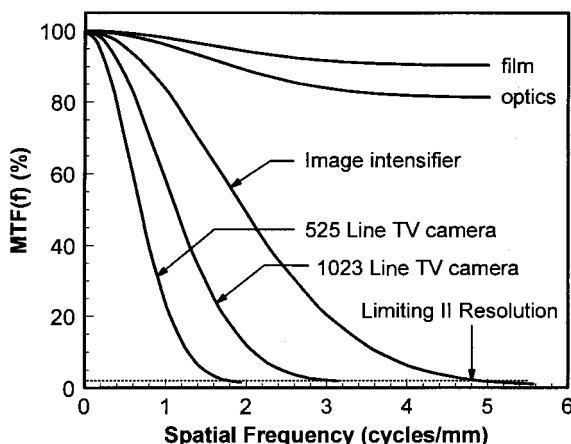


FIGURE 9-14. The modulation transfer function (MTF) of the components in the fluoroscopic imaging chain. The film (for example in a photo-spot camera) has very good resolution, and the optics in the optical distributor also have excellent resolution. The MTF of the image intensifier is shown with a limiting resolution of approximately 4.8 cycles per mm. This MTF is typical for a 23-cm image intensifier mode. The TV camera MTF is the worst in the imaging chain, and limits the MTF of the overall image during live fluoroscopy and videotaped imaging sequences. Image acquisition devices that bypass the video chain (photo spot and cine) enjoy the native resolution of the image intensifier.

TABLE 9-1. TYPICAL LIMITING SPATIAL RESOLUTION IN FLUOROSCOPY

Field of View: cm (inches)	1,023-Line Video: Line Pairs/mm	525-Line Video: Line Pairs/mm
14 (5.5)	2.7	1.4
20 (7.9)	2.0	1.0
27 (10.6)	1.6	0.7
40 (15.7)	1.3	0.5

trast-detail phantoms under fluoroscopic imaging conditions. Contrast resolution is increased when higher exposure rates are used, but the disadvantage is more radiation dose to the patient. The use of exposure rates consistent with the image quality needs of the fluoroscopic examination is an appropriate guiding principle. Fluoroscopic systems with different dose settings (selectable at the console) allow the user flexibility from patient to patient to adjust the compromise between contrast resolution and patient exposure.

Temporal Resolution

The excellent temporal resolution of fluoroscopy is its strength in comparison to radiography, and is its reason for existence. Blurring phenomena that occur in the spatial domain reduce the spatial resolution, and similarly blurring in the time domain can reduce temporal resolution. Blurring in the time domain is typically called image *lag*. Lag implies that a fraction of the image data from one frame carries over into the next frame. In fact, image information from several frames can combine together. Lag is not necessarily bad, as mentioned previously in the section on frame averaging. Video cameras such as the vidicon (SbS₃ target) demonstrate a fair amount of lag. By blurring together several frames of image data, a higher SNR is realized because the contribution of the x-ray photons from the several images are essentially combined into one image. In addition to the lag properties of the video camera, the human eye produces a lag of about 0.2 sec. At 30 frames/sec, this means that approximately six frames are blurred together by the lag of the human eye-brain system. Combining an image with six times the x-ray photons increases the SNR by $\sqrt{6} \approx 2.5$ times, and thus the contrast resolution improves at the expense of temporal resolution. As mentioned above, some systems provide digital frame averaging features that intentionally add lag to the fluoroscopic image.

Whereas some lag is usually beneficial for routine fluoroscopic viewing, for real-time image acquisition of dynamic events such as in digital subtraction angiography, lag is undesirable. With DSA and digital cine, cameras with low-lag performance (e.g., plumbicons or CCD cameras) are used to maintain temporal resolution.

9.7 FLUOROSCOPY SUITES

The clinical application strongly influences the type of fluoroscopic equipment that is needed. Although smaller facilities may use one fluoroscopic system for a wide variety of procedures, larger hospitals have several fluoroscopic suites dedicated to

specific applications, such as GI/GU, cystography, peripheral vascular and cardiac angiography, and neurovascular examinations.

Gastrointestinal Suites

In GI fluoroscopic applications, the radiographic/fluoroscopic room (commonly referred to as an R and F room) consists of a large table that can be rotated from horizontal to vertical to put the patient in a head-down (Trendelenburg) or head-up position (Fig. 9-15A). The II can be either over or under the table, but is usually above the table. The table has a large shroud around it, hiding the x-ray tube or the II and providing radiation shielding. The table usually will have a Bucky tray in it, to be used with an overhead x-ray tube, mounted on a ceiling crane in the room. A spot film device or photo-spot camera often accompanies this table configuration.

Remote Fluoroscopy Rooms

A remote fluoroscopy room is designed for remote operation by the radiologist. The fluoroscopic system is motorized and is controlled by the operator sitting in a lead-glass shielded control booth. In addition to the normal movement of the II (pan, zoom, up, down), remote rooms typically have a motorized compression paddle that is used to remotely palpate the patient. Remote rooms use a reverse geometry, where the x-ray tube is above the table and the image intensifier is below the table. These systems reduce the radiation exposure to the physician operating it, and allow the procedure to be done sitting down without a lead apron on. Remote rooms require more patience and operator experience than conventional rooms.

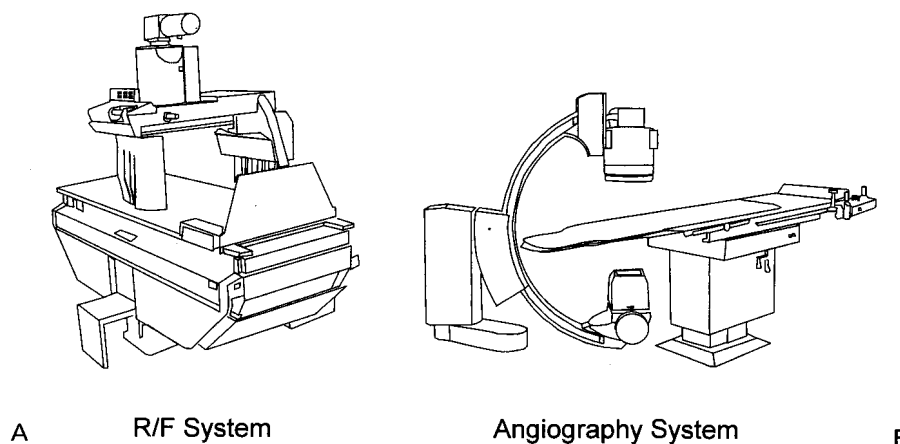


FIGURE 9-15. A: A radiographic/fluoroscopic (R/F) system. R/F systems allow different table angulation to facilitate the movement of contrast agents (typically barium-based) upward or downward in the patient. **B:** Angiography systems typically use a table that does not angulate, but rather floats left to right and top to bottom, which allows easy adjustment of the patient's anatomy with respect to the imaging chain for following a catheter as it is advanced. The x-ray tube (*below*) and image intensifier (*above*) are commonly mounted on a C-arm or U-arm configuration, which allows easy adjustments between the lateral and posteroanterior (PA) imaging positions.

Peripheral Angiography Suites

In the angiographic setting, the table does not rotate but rather “floats”—it allows the patient to be moved from side to side and from head to toe (Fig. 9-15B). The fluoroscopy system (the x-ray tube and II) can also be panned around the stationary patient. The fluoroscopic system is typically mounted on a C-arm or U-arm apparatus, which can be rotated and skewed, and these motions provide considerable flexibility in providing standard posteroanterior (PA) and lateral as well as oblique projections. Because of the routine use of iodinated contrast media, power injectors are often table- or ceiling-mounted in angiographic suites. For peripheral angiography and body work, the image intensifier diameters run from 30 to 40 cm. For neuroangiography rooms, 30-cm image intensifiers are commonly used.

Cardiology Catheterization Suite

The cardiac catheterization suite is very similar to an angiography suite, but typically 23-cm image intensifiers are used. The smaller IIs permit more tilt in the cranial caudal direction, as is typical in cardiac imaging. Cine cameras, either film based or digital, are also mandatory features of the cardiac catheterization suite. Many cardiac catheterization systems are *biplane*.

Biplane Angiographic Systems

Biplane angiographic systems are systems with one table but with two complete imaging chains; there are two generators, two image intensifiers, and two x-ray tubes. Two complete x-ray tube/II systems are used to simultaneously record the anteroposterior and lateral views of a single contrast injection. The simultaneous acquisition of two views allows a reduction of the volume of contrast media that is injected into the patient. During biplane angiography, the x-ray pulse sequence of each plane is staggered so that scattered radiation from one imaging plane is not imaged on the other plane. Biplane suites usually allow one of the imaging chains to be swung out of the way, when single plane operation is desired.

Portable Fluoroscopy—C Arms

Portable fluoroscopy systems are C-arm devices with an x-ray tube placed opposite from the II. Many C-arm systems are 18-cm (7-inch), but 23-cm (9-inch) systems are available as well. Portable C-arm systems are used frequently in the operating room and intensive care units.

9.8 RADIATION DOSE

Patient Dose

The maximal exposure rate permitted in the United States is governed by the Code of Federal Regulations (CFR), and is overseen by the Center for Devices and Radiological Health (CDRH), a branch of the Food and Drug Administration (FDA). This regulation is incumbent upon fluoroscopic equipment manufacturers. The maximal legal entrance exposure rate for normal fluoroscopy to the patient is 10

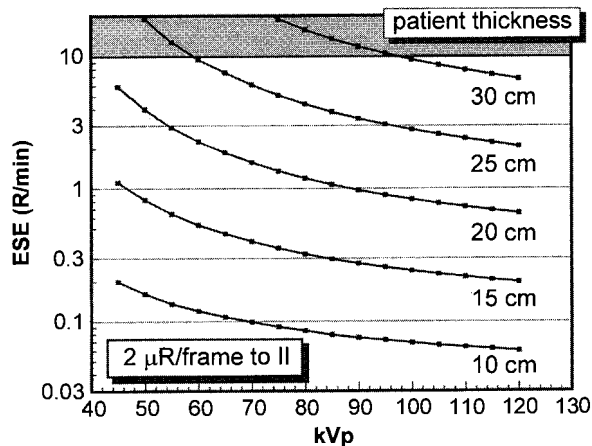


FIGURE 9-16. The entrance skin exposure (ESE) is a function of kVp for different patient thicknesses. These data assume that the fluoroscopic system is adjusted to the 2 mR frame at the input phosphor of the II. An anti-scatter grid with 50% transmission and a postgrid scattered primary ratio of 1.0 is assumed. As the kVp increases, the ESE at all patient thicknesses is reduced. Thick patients imaged at low kV suffer the highest entrance skin exposures. ESE values above 10 R/min are restricted by regulations (gray zone) and the fluoroscopic system has safeguards to prohibit these exposure rates.

R/min. Most state health departments have identical regulations that are imposed on hospitals and private offices. For specially activated fluoroscopy, the maximum exposure rate allowable is 20 R/min. Typical entrance exposure rates for fluoroscopic imaging are about 1 to 2 R/min for thin (10-cm) body parts and 3 to 5 R/min for the average patient. Heavy patients will require 8 to 10 R/min. The entrance skin exposure (ESE) rates for a 2 μ R/frame constant II exposure are shown in Fig. 9-16.

Many methods can be employed during fluoroscopic procedures to reduce the patient dose. Low-dose techniques include heavy x-ray beam filtration (e.g., 0.2 mm copper), aggressive use of low frame rate pulsed fluoroscopy, and use of low-dose (higher kV, lower mA) ABC options. Last-frame-hold features often help in reducing fluoroscopy time. Using the largest field of view suitable for a given clinical study also helps reduce radiation dose to the patient. The most effective way to reduce patient dose during fluoroscopy is to use less fluoroscopy time. The fluoroscopist should learn to use the fluoroscopic image sparingly, and release the fluoroscopy pedal frequently. Dose during fluoroscopy procedures also comes from the associated radiographic acquisitions: DSA sequences in the angiography suite, cine-radiography in the cardiac catheterization laboratory, or the acquisition of photo-spot images in other fluoroscopic procedures. Reducing the number of radiographic images, consistent with the diagnostic requirement of the examination, is also a good way of reducing patient dose.

Dose to Personnel

Occupational exposure of physicians, nurses, technologists, and other personnel who routinely work in fluoroscopic suites can be high. As a rule of thumb, standing 1 m from the patient, the fluoroscopist receives from scattered radiation (on the outside of his or her apron) approximately 1/1,000 of the exposure incident upon the patient. Exposure profiles to the fluoroscopist are shown in Fig. 9-17. The first rule of the fluoroscopy suite is that lead aprons are worn, always, and by everybody in the room when the x-ray beam is on. Portable lead glass shields should be available for additional protection to staff members observing or otherwise participating in the procedure. The dose to personnel is reduced when the dose to the patient is

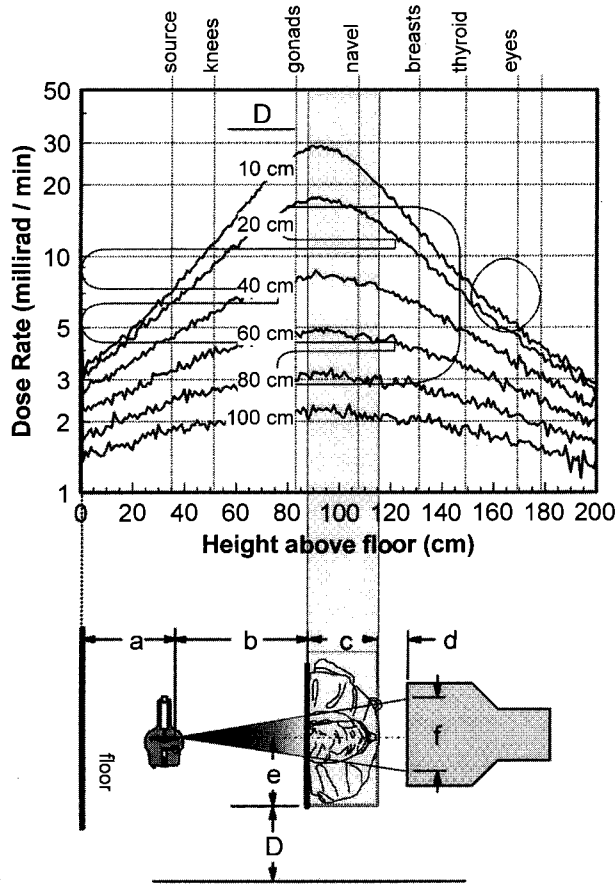


FIGURE 9-17. The scatter field incident upon the radiologist while performing a fluoroscopic procedure. The silhouette of a radiologist of average height, 178 cm (5'10"), is shown overlaid on the graph, and key anatomic levels of the radiologist are indicated at the top. The dose rate (mrad/min) as a function of height above the floor in the fluoroscopic room is shown, for six different distances (D), representing the distance between the edge of the patient and the radiologist (*lower graphic*). A 20-cm-thick patient and a 80-kVp beam were assumed for this calculation. Distances defined in the lower graphic that were used in calculating these data are as follows: $a = 35$ cm, $b = 52$ cm, $c = 28$ cm, $d = 10$ cm, $e = 23$ cm, and $f = 30$ cm.

reduced, so reducing the total fluoroscopy time is beneficial to everyone. Stepping back from the radiation field helps reduce dose to the fluoroscopist (Fig. 9-17). In addition to the lead apron, lead eyeglasses, a thyroid shield, and the use of additional ceiling-mounted lead glass shields can help to reduce the radiation dose to personnel in the fluoroscopy suite. In short, the basic tenets of radiation safety, *time*, *distance*, and *shielding*, apply very well toward dose reduction in the fluoroscopy suite. See Chapter 23 for additional information on radiation protection.

RAP Meters

Roentgen-area product (RAP) meters (also called dose-area product, DAP) are air ionization chambers that are mounted just after the collimator in the imaging chain, and are used to measure the intensity of the x-ray beam striking the patient. The ionization chamber, being mostly air, causes very little attenuation of the x-ray beam. The RAP meter is sensitive not only to the exposure rate and fluoroscopy time, but also to the area of the beam used (hence the term *roentgen-area* product). The RAP meter typically does not record the source-to-patient distance, the magnification mode of the II, or changes in the kV of the x-ray beam, and these all affect radiation dose to the patient. The RAP meter does, however, provide a real-

time estimate of the amount of radiation that the patient has received. The use of a RAP meter is recommended in fluoroscopic rooms where long procedures are common, such as in the angiography suite and the cardiac catheterization laboratory.

SUGGESTED READING

- American Association of Physicist in Medicine Radiation Protection Task Group no. 6. *Managing the use of fluoroscopy in medical institutions*. AAPM report no. 58. Madison, WI: Medical Physics, 1998.
- Code of Federal Regulations 21 Food and Drugs*. 21 CFR §1020.32, Office of the Federal Register National Archives and Records Administration, April 1, 1994.
- Exposure of the U.S. population from diagnostic medical radiation*. NCRP publication no. 100. Bethesda, MD: National Council on Radiation Protection, May 1989.
- Proceedings of the ACR/FDA Workshop on Fluoroscopy, Strategies for Improvement in Performance, Radiation Safety and Control*. American College of Radiology, Reston, VA, October 16–17, 1992.
- Schueler BA. The AAPM/RSNA physics tutorial for residents: general overview of fluoroscopic imaging. *RadioGraphics* 2000;20:1115–1126. Van Lysel MS. The AAPM/RSNA physics tutorial for residents: fluoroscopy: optical coupling and the video system. *RadioGraphics* 2000;20:1769–1786.
- Wagner LK, Archer BR. *Minimizing risks from fluoroscopic x-rays: bioeffects, instrumentation and examination*, 2nd ed. Woodlands, TX: RM Partnership, 1998.
- Wang J, Blackburn TJ. The AAPM/RSNA physics tutorial for residents: x-ray image intensifiers for fluoroscopy. *RadioGraphics* 2000;20:1471–1477.

IMAGE QUALITY

Image quality is a generic concept that applies to all types of images. It applies to medical images, photography, television images, and satellite reconnaissance images. “Quality” is a subjective notion and is dependent on the function of the image. In radiology, the outcome measure of the quality of a radiologic image is its usefulness in determining an accurate diagnosis. More objective measures of image quality, however, are the subject of this chapter. It is important to establish at the outset that the concepts of image quality discussed below are fundamentally and intrinsically related to the diagnostic utility of an image. Large masses can be seen on poor-quality images, and no amount of image fidelity will demonstrate pathology that is too small or faint to be detected. The true test of an imaging system, and of the radiologist that uses it, is the reliable detection and accurate depiction of subtle abnormalities. With diagnostic excellence as the goal, maintaining the highest image fidelity possible is crucial to the practicing radiologist and to his or her imaging facility. While technologists take a quick glance at the images they produce, it is the radiologist who sits down and truly analyzes them. Consequently, understanding the image characteristics that comprise image quality is important so that the radiologist can recognize problems, and articulate their cause, when they do occur.

As visual creatures, humans are quite adroit at visual cognition and description. Most people can look at an image and determine if it is “grainy” or “out of focus.” One of the primary aims of this chapter is to introduce the reader to the terminology used to describe image quality, and to the parameters that are used to measure it. The principal components of image quality are *contrast*, *spatial resolution*, and *noise*. Each of these is discussed in detail below.

10.1 CONTRAST

What is contrast? Figure 10-1 contains two radiographs; the one on the left is uniformly gray. X-rays were taken, film was exposed, but the image remains vacant and useless. The chest radiograph on the right is vivid and useful, and is rich with anatomic detail. The image on the right possesses contrast, the image on the left has none. Contrast is the *difference* in the image gray scale between closely adjacent regions on the image. The human eye craves contrast; it is the pizzazz in the image. The very terms that describe low contrast images, such as “flat” or “washed out,” have negative connotations. The contrast present in a medical image is the result of

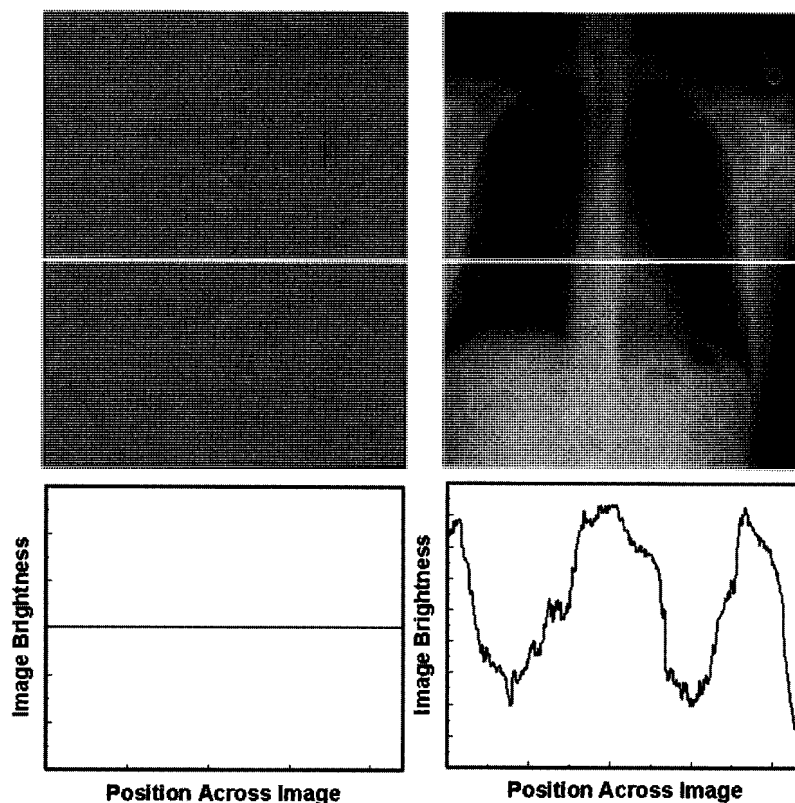


FIGURE 10-1. The homogeneous gray image (*left*) illustrates a radiographic projection void of contrast. The chest radiograph (*right*) demonstrates a wide degree of contrast, as illustrated by the profile of image density below the image. The profile under the left image is flat, demonstrating no contrast.

a number of different steps that occur during image acquisition, processing, and display. Different definitions of contrast occur for each step in the imaging process, and these different mechanisms are described below.

Subject Contrast

Subject contrast is the difference in some aspect of the signal, prior to its being recorded. Subject contrast can be a consequence of a difference in intensity, energy fluence, x-ray energy, phase, radionuclide activity, relaxation characteristics, and so on. While most two-dimensional (2D) medical images are spatial in nature, and subject contrast is mediated through spatially dependent events, time-dependent phenomena can produce spatially dependent subject contrast in some modalities such as in digital subtraction angiography or magnetic resonance imaging. Subject contrast is a result of fundamental mechanisms, usually an interaction between the type of energy (“carrier wave”) used in the modality and the patient’s anatomy or physiology. While subject contrast (like all concepts of contrast) is defined as the difference in some aspect of the signal, the magnitude of the difference (and the amount of con-

trast in the resultant image) can often be affected or adjusted by changing various parameters on the imaging system (turning the knobs). Two examples will illustrate the concept of subject contrast: radiography and magnetic resonance imaging.

Figure 10-2 illustrates a simple situation in projection radiography; a homogeneous beam of (N_0) x-ray photons is incident upon a slab of tissue with linear attenuation coefficient μ and thickness x , and toward the right there is a small bump of tissue of additional thickness z . Although the increased bump of tissue is shown on top of the tissue slab here, more often it may be inside the patient, such as a pulmonary nodule filling in what would be air in the lungs. Under the two different regions, the x-ray fluence (e.g., photons/cm²) exiting the slab of tissue (x-ray scatter is excluded in this discussion) is given by A and B (Fig. 10-2), respectively. The subject contrast (C_s) can be defined here as

$$C_s = \frac{(A - B)}{A} \quad [10-1]$$

and because $A > B$, C_s will be a number between 0.0 and 1.0. The simple equations that define x-ray attenuation are

$$\begin{aligned} A &= N_0 e^{-\mu x} \\ B &= N_0 e^{-\mu(x+z)} \end{aligned} \quad [10-2]$$

Substituting these into the expression for subject contrast (Equation 10-1) and rearranging, it can be shown that:

$$C_s = 1 - e^{-\mu z} \quad [10-3]$$

Consistent with our notion of subject contrast being caused by *differences* in the object, the total thickness of the slab of tissue (x) does not appear in the above expression. Only the difference between the two thicknesses (the difference in thickness = z) remains in the equation. When $z = 0$, the bump goes away, $e^{-\mu z} = 1$, and $C_s = 0$. If z increases in thickness, the subject contrast increases.

The above equation also illustrates a very important aspect of x-ray imaging, pertinent here to radiography and especially mammography. Notice that the linear attenuation coefficient of the tissue slab, μ , shows up in the expression for subject contrast. Just like if we increase z , if we increase μ instead, subject contrast will increase. Figure 10-3A shows the linear attenuation coefficient (μ) for tissue, plotted as a function of x-ray energy. At lower energies, μ increases substantially. For example, changing energies from 40 keV to 20 keV, μ increases by a factor of 3, and C_s increases by nearly a factor of 3. So by reducing the x-ray energy, the subject con-

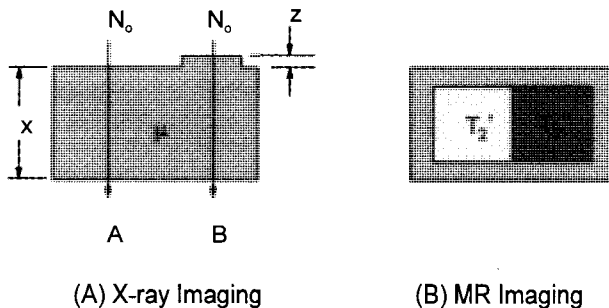


FIGURE 10-2. A: The relationship between the changes in thickness of an object and subject contrast for the x-ray projection imaging example. The subject contrast derived from the measurement of signal A and B is $(1 - e^{-\mu z})$. **B:** The nature of subject contrast in magnetic resonance imaging, where two adjacent regions within the patient have different T2-relaxation constants (T_2' and T_2'').

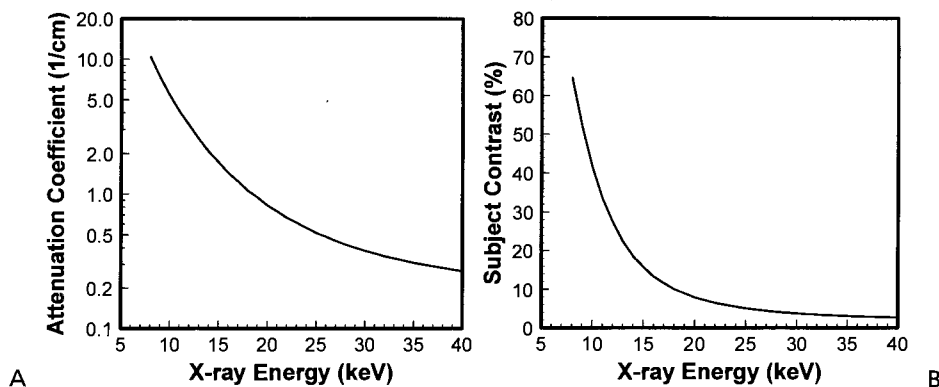


FIGURE 10-3. A: The value of the linear attenuation coefficient (μ in Fig. 10-2A) increases as the x-ray energy decreases. **B:** The subject contrast resulting from the geometry illustrated in Fig. 10-2A is shown as a function of x-ray energy. Lowering the x-ray energy increases the value of μ , which in turn increases the subject contrast of tissue in this example.

trast for the same thickness of tissue increases (Fig. 10-3B). This relatively simple analysis is the entire rationale behind mammography in a nutshell: Breast tissue exhibits low x-ray contrast at conventional radiographic energies (e.g., 35 to 70 keV), but by lowering the x-ray energy substantially as in mammography (20 to 23 keV), higher soft tissue contrast is achieved. This also illustrates the earlier point that changing imaging modality parameters (kV in this example) can often be used to increase subject contrast.

In magnetic resonance imaging (MRI), the signal measured for each pixel in the image is related to nuclear relaxation properties in tissue. Take, for example, two adjacent areas in the patient having different T2 relaxation constants, as shown in

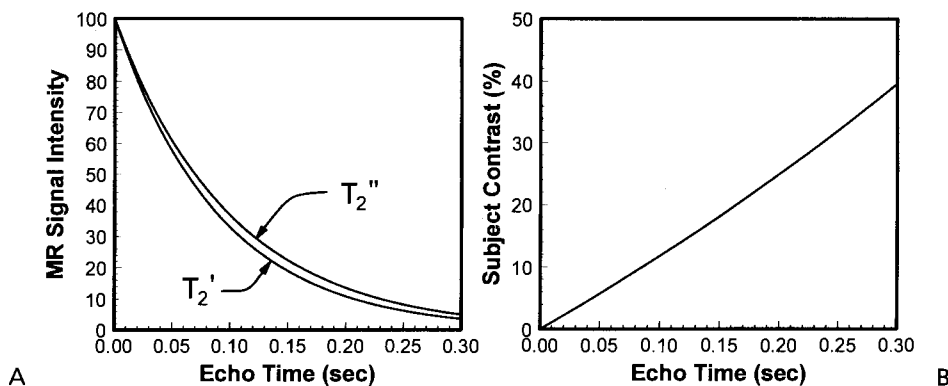


FIGURE 10-4. A: The magnetic resonance imaging (MRI) signal intensity is shown as a function of echo time for two slightly different T2 values ($T_2' = 90$ msec, $T_2'' = 100$ msec), corresponding to the geometry illustrated in Fig. 10-2B. **B:** The subject contrast as demonstrated on MRI increases as a function of echo time, for the two T2 values illustrated (left). By modulating a machine-dependent parameter (TE), the subject contrast resulting from the geometry illustrated in Fig. 10-2B can be increased substantially.

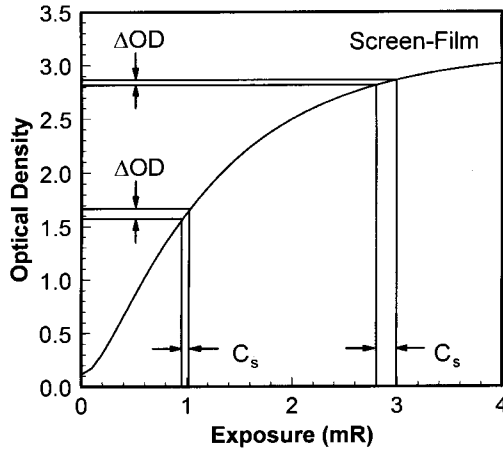


FIGURE 10-5. Due to the nonlinear response of film, the *radiographic contrast* (ΔOD) changes depending on the exposure level. The subject contrast (C_s) shown as vertical lines near 1.0 mR and near 3.0 mR is the same; however, the radiographic contrast differs due to the nonlinear characteristics of the characteristic curve for film.

Fig. 10-4A. A detailed discussion of MRI is unnecessary here, but in simplified terms the signal in MRI is given by

$$\begin{aligned} A &= k e^{-TE/T2'} \\ B &= k e^{-TE/T2''} \end{aligned} \quad [10-4]$$

where TE is the echo time, and k is a constant (which includes proton density and T1 relaxation terms). These equations happen to be identical in form to the x-ray example above, and the subject contrast is given by

$$C_s = 1 - e^{-TE/(T2' - T2'')} \quad [10-5]$$

So if TE is increased, $e^{-TE/(T2' - T2'')}$ decreases, and C_s increases (Fig. 10-4B). Similar to the x-ray example, C_s is dependent on the *difference* between tissue properties ($\Delta T2 = T2' - T2''$), not on the T2 values per se, and adjusting a machine dependent parameter (TE) changes subject contrast. MRI is an example of where a time-dependent property of tissue (T2 relaxation) is used to modulate subject contrast spatially.

Detector Contrast

An imaging detector receives the incident energy striking it, and produces a signal in response to that input energy. Because small changes in input energy striking the detector (i.e., subject contrast) can result in concomitant changes in signal (image contrast), the detector's characteristics play an important role in producing contrast in the final image. Detector contrast, C_d , is determined principally by how the detector "maps" detected energy into the output signal. Different detectors can convert input energy to output signal with different efficiency, and in the process amplify contrast. A diagram that shows how a detector maps input energy to output signal is called a *characteristic curve*, and one is shown for a screen-film system in Fig. 10-5. Evident in Fig. 10-5 is the nonlinear relationship between the input exposure and the output optical density (OD), which is discussed in detail in Chapter 6. Two characteristic curves for digital imaging systems are shown in Fig. 10-6; one curve is linear and the other logarithmic. The output of a digital imaging sys-

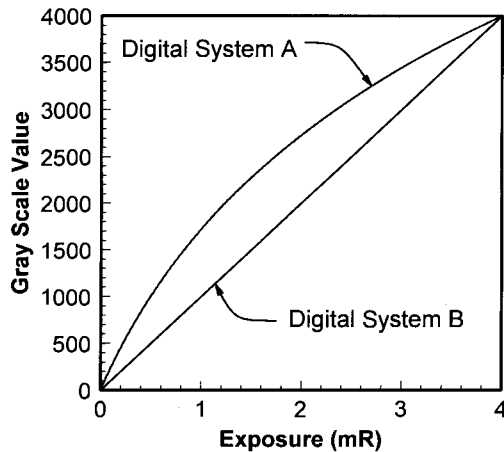


FIGURE 10-6. The characteristic curves for two different digital imaging systems. **A:** A logarithmic response, typical of computed radiography (CR) and film digitizers. **B:** A linear response, typical of image intensifier-television-digitizer and flat panel x-ray imaging systems.

tem is a digital image, and the unit that conveys the light and dark properties for each pixel is called a *gray scale value*, or sometimes a *digital number*. Many digital systems demonstrate linear characteristic curves (e.g., thin film transistor systems), while some have an intentional logarithmic response (computed radiography and film digitizers).

The detector contrast is the slope of the characteristic curve. For screen-film systems the exposure axis is usually plotted using a logarithmic scale, and the slope at any point along the curve is called “gamma” and is given the symbol γ , such that:

$$\gamma = [OD_A - OD_B] / [\text{Log}(A) - \text{Log}(B)] \quad [10-6]$$

For nonlinear detectors such as screen-film, the slope (γ) of the characteristic curve changes depending on exposure level (Fig. 10-7). If the characteristic curve is linear, then the slope will be constant (Fig. 10-7).

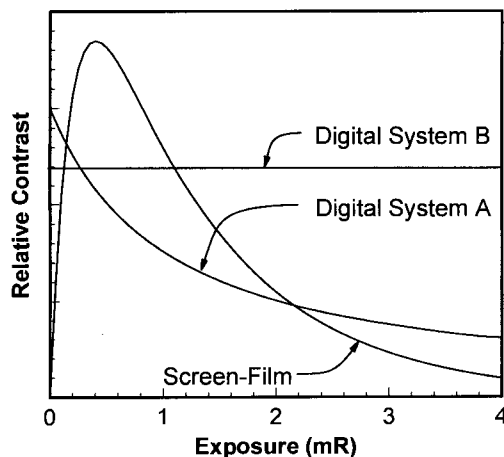


FIGURE 10-7. The contrast properties of detector system are related to the slope of the characteristic curve of the system. The slopes of the characteristic curve of a typical screen-film system (Fig. 10-5) and of two digital systems (Fig. 10-6) are illustrated here. The screen-film system starts out with low detector contrast at low exposure, and with increasing exposure it rapidly increases to a maximum value at around 0.7 mR, corresponding to the steepest point on the H and D curve shown in Fig. 10-5. The slope of the log curve for digital system A in Fig. 10-6 starts high and decreases steadily with increasing exposure levels. The slope of a straight line (digital system B) in Fig. 10-6 is a constant, illustrated as the *horizontal line* on this figure.

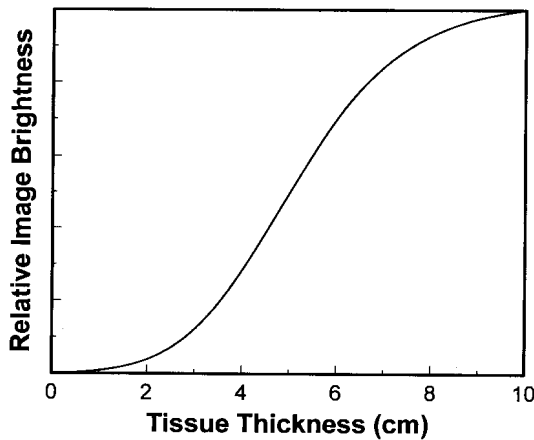


FIGURE 10-8. The relative image brightness (which is related to contrast) as a function of patient thickness is dependent on the subject contrast, the detector contrast, and the overall brightness of the light box. The combined influence of all the image formation steps yields a sigmoidal relationship between image brightness reaching the radiologist's eyes and patient thickness.

Radiographic Contrast (Screen-Film)

There is sufficient difference between digital imaging systems and screen-film radiography that the output measure of contrast in the image is defined separately. For screen-film radiography and mammography, the analog film is the output and the contrast that is seen on the film is called *radiographic contrast*:

$$\text{Radiographic Contrast} = OD_A - OD_B \quad [10-7]$$

where A and B refer to adjacent regions on the radiograph. While a more complete discussion of optical density is found in Chapter 6, it is worth mentioning here that radiographic films are placed over light boxes for display. In this process, the number of light photons hitting the radiologist's eyes is proportional to $B 10^{-OD}$, where B is the brightness of the light box. In radiography, the contrast cannot be adjusted or enhanced on the analog film. So how does the final light signal that reaches the radiologist's eyes depend on patient thickness? The mathematical relationships relating to subject contrast, detector contrast, and $B 10^{-OD}$ were combined to demonstrate the relationship between brightness on the radiograph and tissue thickness, shown in Fig. 10-8.

Digital Image Contrast (Contrast-to-Noise Ratio)

Once a digital image is acquired, for many types of detector systems, a series of processing steps is performed automatically as part of the acquisition software. The details are dependent on the modality and detector technology, but one common type of processing is the subtraction of a constant number (call this number k) from the image. Let the average density of a small square region on a digital image be denoted as A , but after the processing this becomes $A - k$. The average value of an adjacent region can be denoted $B - k$, and assume that $A > B$. If we apply previous notions of contrast to these values, such as for subject contrast (Equation 10-1), we immediately run into problems when the processing is considered: $\text{Contrast} = ([A - k] - [B - k]) / [A - k] = (A - B) / (A - k)$. Notice that if $k = A/2$, contrast is doubled, and if $k = A$, contrast would be infinite (division by zero). If k is negative, contrast is reduced. Thus, depending on the somewhat arbitrary choice of k , contrast

can be radically changed. Therefore, the whole notion of contrast on a digital image has to be rethought.

A more meaningful and frequently used measure in assessing digital images, related to contrast, is the *contrast-to-noise ratio* (CNR):

$$\text{CNR} = (A - B)/\sigma \quad [10-8]$$

where the noise in the image is designated as σ . Noise will be discussed in more detail below, but a couple of observations are warranted here. The CNR is not dependent on k , and that is good since the selection of k (an offset value) can be arbitrary. More importantly, because of the ability to postprocess digital images (unlike analog radiographic images), the CNR is a more relevant description of the contrast potential in the image than is contrast itself.

Displayed Contrast (Digital Images)

One of the most powerful attributes of digital images is that their displayed appearance can be easily changed. Figure 10-9 illustrates a digital chest radiograph displayed using two different display curves. The graph in Fig. 10-9 illustrates a histogram of the image, and the two curves A and B illustrate the actual display curves used to produce the two images. The graph in Fig. 10-9 shows the *look-up table* used to convert ("map") the digital image data residing in the com-

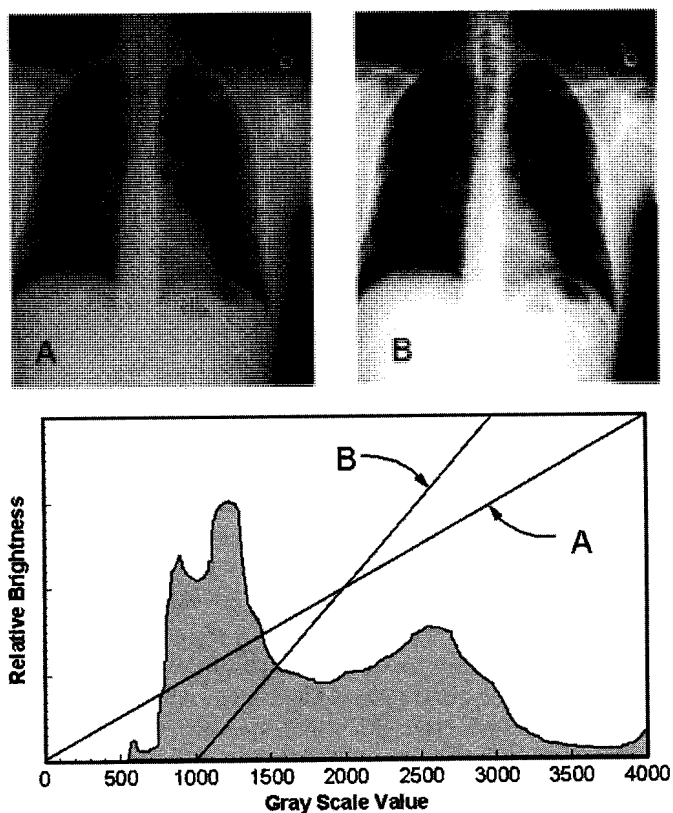


FIGURE 10-9. One of the major benefits of digital images is that displayed contrast can be manipulated, usually in real time, by changing the way in which the gray-scale values are converted into brightness on the display hardware. Here the same chest radiograph is displayed with two different look-up tables (lines A and B on the lower graph). The graph below the two images illustrates a histogram of the gray-scale value in this image. The steeper display look-up table (B) yields a higher contrast image.

puter to the display hardware itself. The appearance of the image can be changed by manipulating the look-up table (i.e., changing the window and level settings, as discussed in Chapter 11), and this process can be done using real-time visual feedback from the display.

A 12-bit digital image has gray scale values running from 0 to 4095, but often the exposure conditions result in a fair amount of the dynamic range not being used. For example, from the image histogram in Fig. 10-9, no gray scale values below 550 are found in that chest radiograph. Look-up table A wastes some fraction (~15%) of the display system's relative brightness by mapping gray scale values from 0 to 550, which do not exist in the image. The slope of the look-up table is related to the displayed image contrast, and there is an analogy here between display look-up tables and detector characteristic curves, discussed previously. Digital images are affected by both curves.

10.2 SPATIAL RESOLUTION

A two-dimensional image really has three dimensions: height, width, and gray scale. The height and width dimensions are spatial (usually), and have units such as millimeters. Spatial resolution is a property that describes the ability of an imaging system to accurately depict objects in the two spatial dimensions of the image. Spatial resolution is sometimes referred to simply as the *resolution*. The classic notion of spatial resolution is the ability of an image system to distinctly depict two objects as they become smaller and closer together. The closer together they are, with the image still showing them as separate objects, the better the spatial resolution. At some point, the two objects become so close that they appear as one, and at this point spatial resolution is lost.

Spatial Domain: The Point Spread Function

The *spatial domain* simply refers to the two spatial dimensions of an image, for instance its width (x-dimension) and length (y-dimension). One conceptual way of understanding (and measuring) the spatial resolution of a detector system in the spatial domain is to stimulate the detector system with a single point input, and then observe how it responds. The image produced from a single point stimulus to a detector is called a point response function or a *point spread function* (PSF). For example, a tiny hole can be drilled into a sheet of lead and used to expose an x-ray detector to a very narrow beam of x-rays. A point source of radioactivity can be used in nuclear imaging. For modalities that are tomographic, a point spread function can be produced by placing a line stimulus through (and perpendicular to) the slice plane. A very thin metal wire is used in computed tomography (CT), a monofilament is used in ultrasound, and a small hole filled with water can be used in MRI.

Figure 10-10 illustrates both a point stimulus (Fig. 10-10A) and resulting point spread functions. Figure 10-10B is an isotropic PSF, demonstrating a system that spreads or blurs evenly in all radial directions (e.g., a screen-film system). A non-isotropic PSF is shown in Fig. 10-10C, and nonisotropic PSFs are found with fluoroscopy and tomographic systems. Figure 10-10D illustrates how, for a tomo-

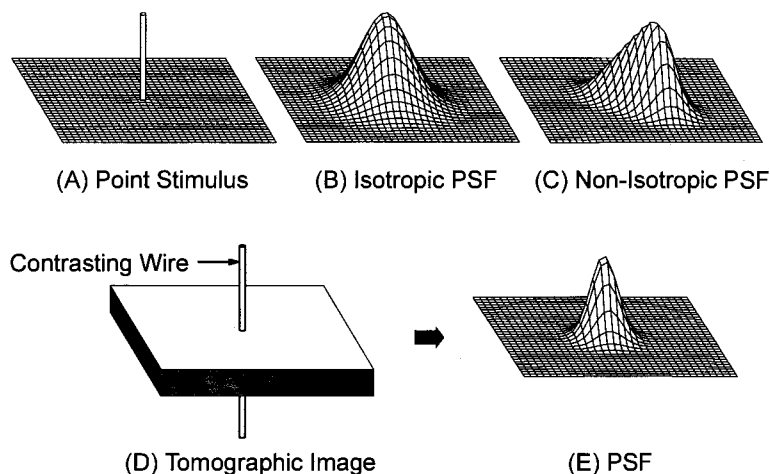


FIGURE 10-10. **A:** An isometric display of a point stimulus input to an imaging system. **B:** The response of an imaging system to a point stimulus is the *point spread function*. The point spread function illustrated here is *isotropic*, meaning that it is rotationally symmetric. **C:** A nonisotropic point spread function, typical of fluoroscopic and tomographic imaging systems. **D:** A point stimulus to a tomographic imaging system (CT, MRI, ultrasound) can be a long thin contrasting wire running perpendicular to the tomographic plane. **E:** The point spread function resulting from the contrasting wire.

graphic image, a line stimulus is used to produce a point stimulus in plane, resulting in a PSF in the tomographic image (Fig. 10-10E).

If the PSF is measured at many different locations and is the same regardless of location, the imaging system is said to be *stationary*. If the PSF changes as a function of position in the image, the detector system is considered *nonstationary*. These concepts are shown in Fig. 10-11. Most imaging systems in radiology fall somewhere between these extremes, but for convenience are considered approximately stationary.

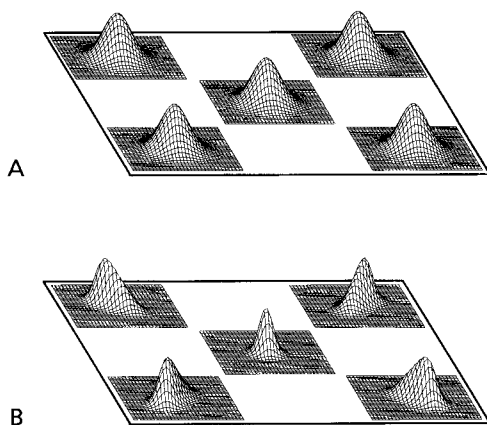


FIGURE 10-11. **A:** A stationary image is one in which the point spread function (PSF) is the same over the entire field of view. Here an isotropic point spread function is shown; however, an imaging system with a nonisotropic PSF is still stationary as long as the PSF is constant over the field of view. **B:** A nonstationary system demonstrates different PSFs, depending on the location in the image.

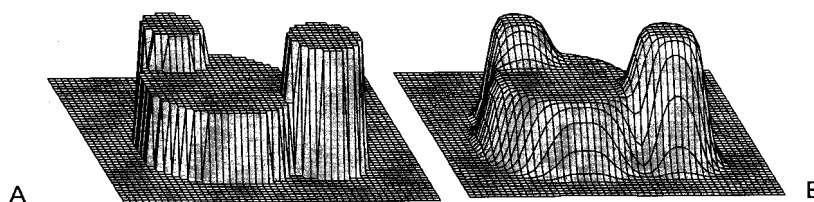


FIGURE 10-12. A: An isometric plot of a simple image of three circles of varying contrast (different heights on this display). **B:** The same image is shown after the blurring influence of an imperfect imaging system occurs. The input to the imaging system has sharp, crisp edges; the imperfect resolution properties of the imaging system smooths out (blurs) edges in the image.

The PSF describes the blurring properties of an imaging system. If you put in a point stimulus to the imaging system, you get back the PSF. But an image is just a large collection of individual points, and if you expose the imaging system to, for example, a medical image, the PSF acts to blur each and every one of the millions of point inputs that comprise the image. Figure 10-12 illustrates (using isometric display) an image consisting of three circular regions of various intensity (intensity shows up as height in the isometric display), before and after the blurring influence of the imaging system. The process of breaking up an input image into its constituent point stimuli, individually blurring each point using the PSF of the imaging system, and then adding up the net result is a mathematical operation called *convolution*. Convolution describes mathematically what happens to the signal physically. Convolution is discussed in other chapters and is a recurrent theme in medical imaging.

Physical Mechanisms of Blurring

There are many different mechanisms in radiologic imaging that cause blurring. The spatial resolution of an image produced by any optical device (e.g., a 35-mm camera) can be reduced by defocusing. When an x-ray strikes an intensifying screen, it produces a burst of light photons that propagate by optical diffusion through the screen matrix. For thicker screens, the diffusion path toward the screen's surface is longer and more lateral diffusion occurs, which results in a broader spot of light reaching the surface of the screen and consequently more blurring (Fig. 10-13A).

Motion is a source of blurring in all imaging modalities. Motion blurring can occur if the patient moves during image acquisition, or if the imaging system moves. Cardiac motion is involuntary, and some patients (e.g., pediatric) have difficulty remaining still and holding their breath during an imaging study. The best way to reduce motion blurring is to reduce the image acquisition time.

In tomographic imaging modalities such as CT, MRI, and ultrasound, slice thickness plays a role in degrading spatial resolution. For anatomic structures that are not exactly perpendicular to the tomographic plane, the edges of the structure will be blurred proportional to the slice thickness and angle of the structure with respect to the tomographic plane (Fig. 10-13B).

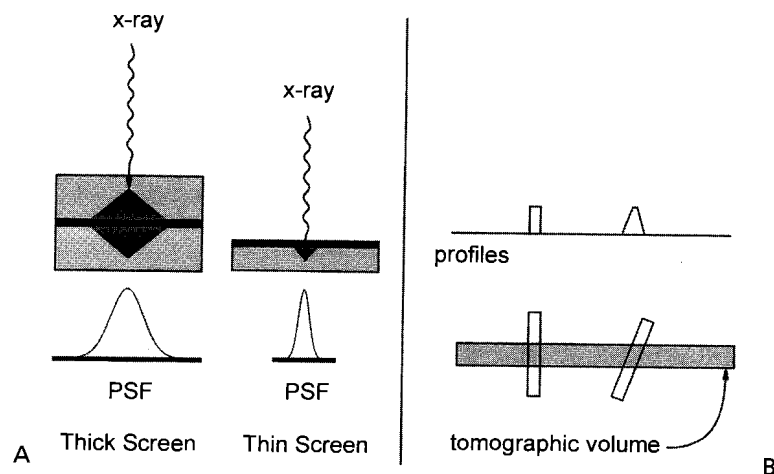


FIGURE 10-13. Some physical mechanisms that result in imaging system blurring. **A:** A thicker intensifying screen absorbs more of the incident x-ray beam; however, the light released within the intensifying screen has a longer diffusion path to the film emulsion, resulting in increased blurring (wider point spread function). **B:** In tomographic imaging modalities, objects that run perpendicular to the tomographic plane will demonstrate sharp edges as seen in the upper profile, whereas objects that intersect the tomographic plane at a nonnormal angle will demonstrate edges with reduced sharpness, as seen on the profile on the right.

Image Magnification

As discussed in Chapter 6, magnification of an object onto the imaging plane occurs in x-ray imaging. Magnification interacts with a finite-sized x-ray tube focal spot to degrade spatial resolution, with greater degradation occurring at higher magnifications. Paradoxically, magnification is also used in some settings to *improve* spatial resolution. Magnification is unavoidable in projection radiography, because x-rays diverge (in both dimensions) as they leave the x-ray tube. This geometry is illustrated in Fig. 10-14. If a flat, thin object is imaged in contact with the detec-

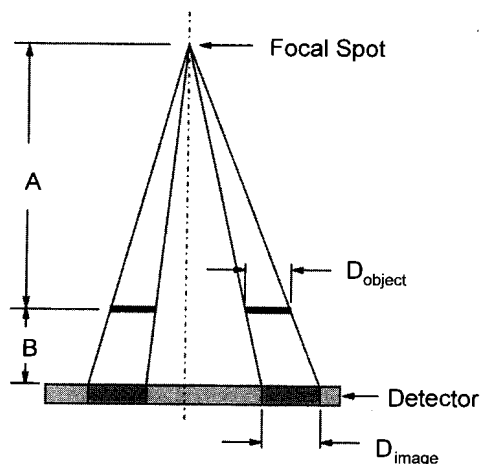


FIGURE 10-14. The geometry for calculating object magnification. Because x-rays diverge from the focal spot, the width of an object (D_{object}) will be increased on the image (D_{image}), depending on the magnification. For a thin object in contact with the image receptor ($B = 0$), magnification becomes 1.0 ($D_{\text{image}} = D_{\text{object}}$).

tor, it will appear life size (magnification = 1.0). However, if the object is not right up against the detector, it will be imaged with some amount of magnification (magnification > 1.0). The relationship between object and image is

$$\text{Object Size} \times M_{\text{object}} = \text{Image Size} \quad [10-9]$$

where M_{object} is the object's magnification factor.

Because x-ray beams are divergent and not convergent, *minification* (magnification factors less than 1.0) cannot occur in projection x-ray imaging. Radiographic images are sometimes used to measure object sizes, for example in sizing vessels for catheter selection or in determining the size of a tumor or other object on the image. Knowing the magnification factor, the above equation can be inverted and object size can be calculated from its dimensions measured on the image. The magnification factor, M_{object} , is calculated (Figure 10-14) as

$$M_{\text{object}} = \frac{A + B}{A} = \frac{\text{Source to Image Distance (SID)}}{\text{Source to Object Distance (SOD)}} \quad [10-10]$$

Notice that this geometry defines the magnification of an *object in the patient*, where that object is between the source and the detector. If the object is up against the detector, $B = 0$, and $M_{\text{object}} = 1.0$.

Magnification influences the amount of blurring that a finite-sized x-ray focal spot causes (Figure 10-15), and the geometry for calculating this is different from that described above and therefore a different expression is required:

$$M_{\text{source}} = \frac{B}{A} \quad [10-11]$$

In general-purpose radiographic imaging, relatively large (~0.6 to 1.5 mm) x-ray focal spots are used, and increased magnification will cause increased blurring of the anatomic detail in the patient, as illustrated in Fig. 10-15. In this geometry, the finite-sized focal spot blurs the edges of the objects and reduces their resolution. The blur severity is determined by the projected width of the focal spot, D_{image} of the focal spot (see Fig. 10-15), also called the *penumbra*. However, if a very small focal spot is used, magnification can *improve* spatial resolution. This is because in

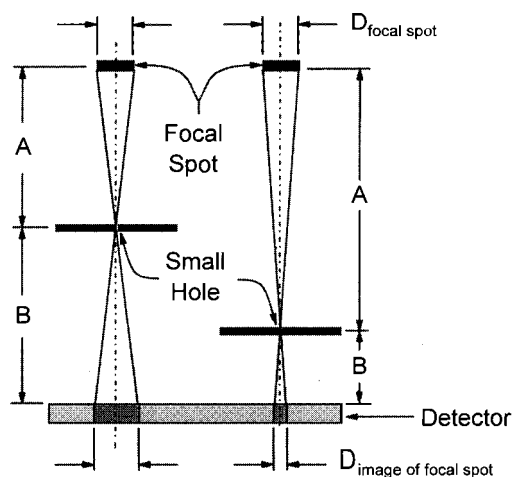


FIGURE 10-15. The geometry associated with x-ray focal spot magnification. An image of the focal spot can be produced using a pinhole, as illustrated. The image of the x-ray source becomes more magnified as the pinhole is moved toward the source and away from the detector (*left*). When the pinhole is halfway between source and detector, the focal spot will be imaged life size (magnification = 1.0). Increasing magnification of an object causes increased blurring due to the focal spot.

some clinical applications, the blurring of the image receptor is the dominant factor in degrading spatial resolution. Examples are in mammography, when very small microcalcifications need to be demonstrated occasionally ("mag views"), or in digital subtraction angiography, where the pixel size limits spatial resolution. In these situations, purposeful magnification can be used with very small focal spots (e.g., 0.10 mm in mammography, 0.30 mm in neuroangiography) to achieve improved spatial resolution. With magnification radiography, the object's x-ray shadow is cast over a larger region of the detector surface, spreading out structures so that they are not blurred together by the detector point spread function (which is not affected by the magnification factor). Magnification radiography only works over a selected range of magnification factors, and at some point the focal spot, however small, will still cause a loss of resolution with increasing magnification. If the focal spot were a *true* point source, increased magnification would always improve spatial resolution based on simple geometry, but other factors would mitigate this in practice (e.g., field of view and motion artifacts). Purposeful magnification radiography is used *only* to overcome the resolution limitations of the detector.

Other Spread Functions

The point spread function describes the response of an imaging system to a point stimulus, and it is a very thorough description of the system's spatial resolution properties. For some imaging systems, however, it is difficult to experimentally measure a PSF. For example, to measure the PSF of a screen-film system, a very small hole (0.010-mm diameter) needs to be aligned with the focal spot 1,000 mm away, a very difficult task. Under these circumstances, other spread functions become useful. The line spread function (LSF) describes the response of an imaging system to a linear stimulus (Fig. 10-16). To determine the LSF, a slit image is acquired and a 90-degree profile across the slit is measured. The LSF can be thought of as a linear collection of a large number of PSFs. The LSF is slightly easier to measure experimentally, because the linear slit that is used need only be aligned with the focal spot in one dimension (whereas the hole used to produce the PSF needs to be aligned in two dimensions).

For a stationary imaging system with an isotropic PSF (Fig. 10-10), a single LSF determination is adequate to fully describe the imaging system's behavior. However, if the PSF is nonisotropic, the LSF needs to be measured with the line stimulus positioned at different angles with respect to the imaging system for a full understanding

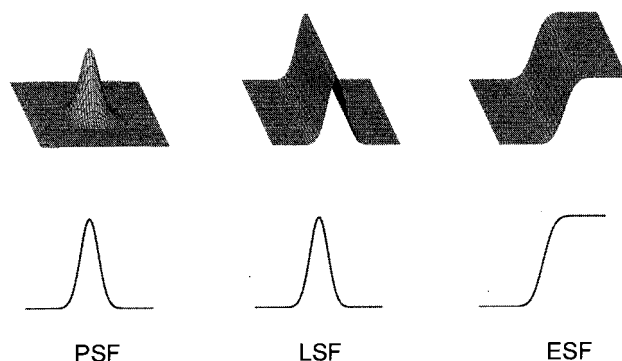


FIGURE 10-16. The point spread function (PSF), line spread function (LSF), and edge spread function (ESF) are shown isometrically (top) and in profile (bottom).

of the resolution performance of that system. With such systems (e.g., conventional fluoroscopy systems), the LSF is often determined in both the horizontal and vertical axes of the imaging system. The LSF function is, in fact, more commonly measured than is the PSF, for reasons that will be described later in this chapter.

The edge spread function (ESF) is sometimes measured instead of the LSF (Fig. 10-16). Rather than a hole (for the PSF) or a line (for the LSF), only a sharp edge is needed to measure the ESF. The ESF is measured in situations where various influences to the imaging system are dependent on the area exposed. For example, the spatial properties of scattered x-ray radiation are often measured using edges. Very little scatter is produced with a small linear stimulus (the LSF), but with the ESF measurement half the field is exposed, which is sufficient to produce a measurable quantity of scattered radiation. For systems that have a large amount of optical light scattering ("veiling glare"), such as fluoroscopic imaging systems, the ESF is useful in characterizing this phenomenon as well.

The Frequency Domain

The PSF and other spread functions are apt descriptions of the resolution properties of an imaging system in the spatial domain. Another useful way to express the resolution of an imaging system is to make use of the spatial *frequency domain*. Most readers are familiar with sound waves and temporal frequency. The amplitude of sound waves vary as a function of time (measured in seconds), and temporal frequency is measured in units of cycles/sec (sec^{-1}), also known as hertz. For example, the note middle A on a piano corresponds to 440 cycles/second. If the peaks and troughs of a sound wave are separated by shorter periods of time, the wave is of higher frequency. Similarly, for objects on an image that are separated by shorter distances (measured in millimeters), these objects correspond to high spatial frequencies (cycles/mm).

Figure 10-17 illustrates a (spatial domain) sine wave spanning a total distance of 2Δ , where Δ is measured in millimeters. If $\Delta = 0.5$ mm, for example, then the single cycle of the sine wave shown in Fig. 10-17 would be 1.0 mm across, and this

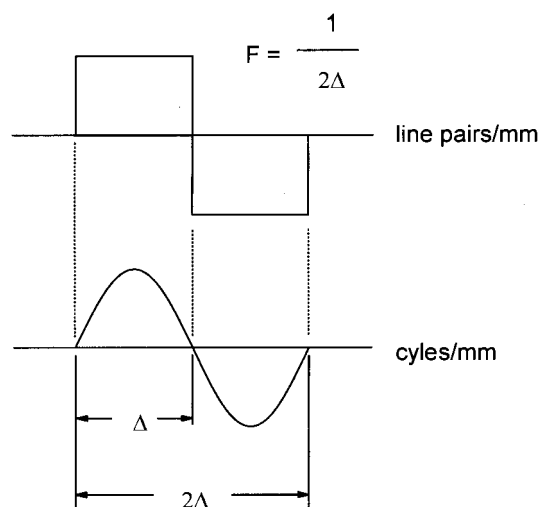


FIGURE 10-17. The concept of *spatial frequency*. A single sine wave (bottom) with the width of one-half of the sine wave, which is equal to a distance Δ . The complete width of the sine wave (2Δ) corresponds to one *cycle*. With Δ measured in millimeters, the corresponding spatial frequency is $F = 1/2\Delta$. Smaller objects (small Δ) correspond to higher spatial frequencies, and larger objects (large Δ) correspond to lower spatial frequencies. The square wave (top) is a simplification of the sine wave, and the square wave shown corresponds to a single line pair.

would correspond to 1 cycle/mm. If $\Delta = 0.10$ mm, then one complete cycle would occur every $2\Delta = 0.20$ mm, and thus in the distance of 1 mm, five cycles would be seen, corresponding to a spatial frequency of 5.0 cycles/mm. The relationship between the distance spanned by one-half cycle of a sine wave, Δ , and the spatial frequency F , is given by:

$$F = \frac{1}{2\Delta} \quad [10-12]$$

In addition to the sine wave shown in Fig. 10-17, a square wave is shown above it. Whereas the spatial frequency domain technically refers to the frequency of sine wave-shaped objects, it is a common simplification conceptually to think of the sine wave as a square wave. The square wave is simply a pattern of alternating density stripes in the image. With the square wave, each cycle becomes a *line pair*—the bright stripe and its neighboring dark stripe. Thus, the units of spatial frequency are sometimes expressed as *line pairs/mm* (lp/mm), instead of *cycles/mm*. A square object of width Δ can be loosely thought of as corresponding to the spatial frequency given by Equation 10-12. So, objects that are 50 μm across (0.050 mm) correspond to 10 line pairs/mm, and objects that are 0.250 mm correspond to 2 line pairs/mm, and so on.

Spatial frequency is just a different way of thinking about object size. There is nothing more complicated about the concept of spatial frequency than that. Low spatial frequencies correspond to larger objects in the image, and higher spatial frequencies correspond to smaller objects. If you know the size of an object (Δ), you can convert it to spatial frequency ($F = 1/2\Delta$), and if you know the spatial frequency (F), you can convert it to object size ($\Delta = 1/2F$). With an understanding of spatial frequency in hand, how it relates to the spatial resolution of an imaging system can now be discussed.

The Modulation Transfer Function: MTF(f)

Let's start with a series of sine waves of different spatial frequencies, as shown in Fig. 10-18. The six sine waves shown in Fig. 10-18 have spatial frequencies of 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 cycles/mm, and these correspond (via Equation 10-12) to object sizes of 1.0, 0.50, 0.333, 0.25, 0.40, and 0.167 mm, respectively. Each sine

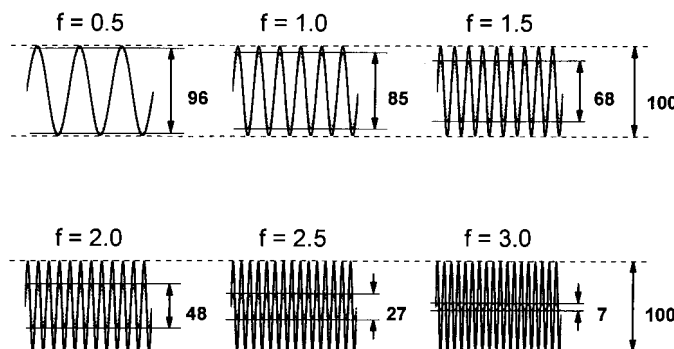


FIGURE 10-18. A series of sine waves of differing spatial frequencies (e.g., $f = 0.5$ cycles/mm). The contrast (difference between peaks and valleys) of the input sine waves (solid lines) is 100%. After detection, the output sine waves (dotted lines) have reduced amplitude (the output sine wave amplitude is indicated for each sine wave). The reduction is more drastic for higher frequency sine waves.

wave serves as an input to a hypothetical imaging system, and the amplitude of each input sine wave corresponds to 100 units. The amplitude here is a measure of the image density (e.g., optical density for film, or gray scale units for a digital image) between the peaks and valleys of the sine wave. Each of the input sine waves is blurred by the point spread function of the imaging system, and the resulting blurred response to each sine wave (the output of the imaging system) is shown in Fig. 10-18 as dotted lines. Notice that as the spatial frequency increases, the blurring causes a greater reduction in the output amplitude of the sine wave.

The amplitude of the sine wave is really just the contrast between the peaks and valleys. All six sine waves in Fig. 10-18 have the same input contrast to the hypothetical imaging system (100 units), but the output contrast was altered by the blurring influence of the point spread function. The output contrast is lower for higher spatial frequencies (i.e., smaller objects), and is identified on Fig. 10-18 by two horizontal lines for each sine wave. The modulation is essentially the output contrast normalized by the input contrast. The modulation transfer function (MTF) of an imaging system is a plot of the imaging system's modulation versus spatial frequency. In Fig. 10-19, the output modulation for each of the sine waves shown in Fig. 10-18 is plotted on the y-axis, and the frequency of the corresponding sine wave is the x-axis value.

The MTF of an image system, like that shown in Fig. 10-19, is a very complete description of the resolution properties of an imaging system. The MTF illustrates the fraction (or percentage) of an object's contrast that is recorded by the imaging system, as a function of the size (i.e., spatial frequency) of the object. To the reader previously unfamiliar with the concept of the MTF, it is fair to ask why imaging scientists prefer to use the MTF to discuss the spatial resolution of an imaging system, over the easier-to-understand spread function description discussed previously. A partial answer is seen in Fig. 10-20. Many imaging systems are really imaging chains, where the image passes through many different intermediate steps from the input to the output of the system (fluoroscopy systems are a good example). To understand the role of each component in the imaging chain, the MTF is measured separately for each component (MTF curves A, B, and C in Fig. 10-20). The total system MTF at any frequency is the *product* of all the subcomponent MTF curves.

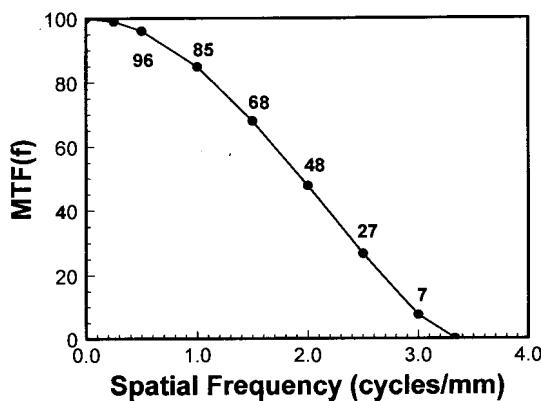


FIGURE 10-19. The output amplitude of the sine waves illustrated in Fig. 10-18 is plotted here on the y-axis, and the spatial frequency is plotted on the x-axis. The *modulation transfer function* is a plot that demonstrates the resolution capabilities of an imaging system (signal modulation) as a function of spatial frequency. Since low spatial frequencies correspond to large objects, and high spatial frequencies correspond to smaller objects, the MTF is just a description of how an imaging system responds depending on the size of the stimulus.

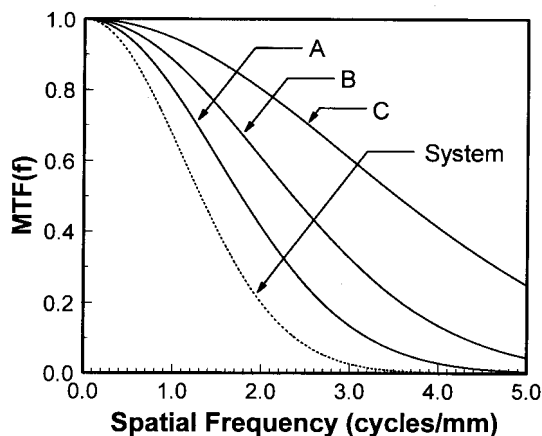


FIGURE 10-20. An imaging system is often made of a series of components, and the MTFs for components A, B, and C are illustrated. The system MTF is calculated by multiplying the (three) individual subcomponent MTFs together, at each spatial frequency. Because MTF values are less than 1.0, the system MTF will always be lower than the MTF of the worst subcomponent. As with a real chain, an imaging chain is as strong as its weakest link.

The LSF, the MTF, and the Fourier Transform

Whereas the MTF is best conceptualized using sine waves, in practice it is measured differently. As mentioned above, the line spread function (LSF) is measured by stimulating the imaging system to a line. For an x-ray detector, a very narrow lead slit is used to collimate the x-ray beam down to a very thin line. For tomographic imaging systems, a very thin sheet of contrasting material can be used to create an LSF (e.g., a 0.1-mm sheet of aluminum for CT). Once the LSF is measured, the MTF can be computed directly from it using the Fourier transform (FT):

$$\text{MTF}(f) = |\text{FT}\{\text{LSF}(x)\}| \quad [10-13]$$

The Fourier transform is an integral calculus operation, but in practice the Fourier transform is performed using a computer subroutine. For those interested in the details of the Fourier transform, a more detailed discussion is given in a text by Bracewell (1978). More important than the mathematical details, the trends

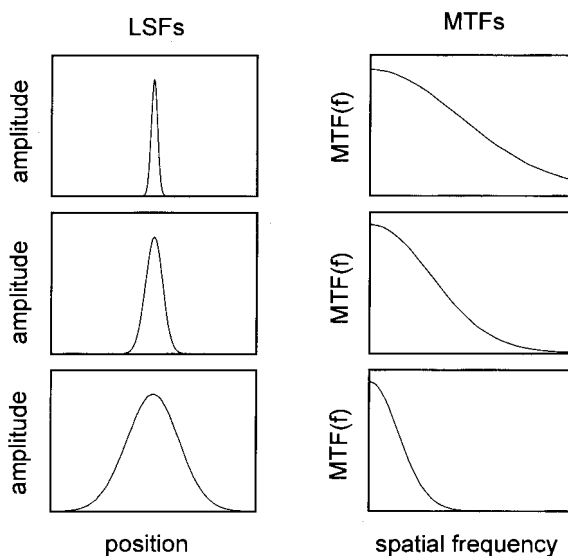


FIGURE 10-21. The MTF is typically calculated from a measurement of the line spread function (LSF). As the line spread function gets broader (left column, top to bottom), the corresponding MTFs plummet to lower MTF values at the same spatial frequency, and the cutoff frequency (where the MTF curve meets the x-axis) is also reduced. The best LSF-MTF pair is at the top, and the worst LSF-MTF pair is at the bottom.

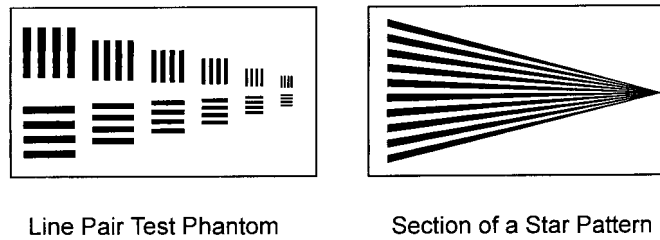


FIGURE 10-22. While the MTF is the best description of the resolution properties of an imaging system, for routine quality assurance purposes easier-to-measure estimates of spatial resolution suffice. The limiting resolution of an imaging system can reliably be estimated using either line pair or star pattern phantoms.

between the LSF and the MTF should be understood. Figure 10-21 illustrates three LSFs, getting progressively worse toward the bottom of the figure. On the right column of Fig. 10-21, the MTFs corresponding to the LSFs are shown. As the LSF gets broader (poorer resolution), the MTF curves dive down toward zero modulation more rapidly.

Field Measurements Using Resolution Templates

Spatial resolution is something that should be monitored on a routine basis for many imaging modalities. However, measuring the LSF or the MTF is usually too detailed for routine quality assurance purposes. For a “quick and dirty” evaluation of spatial resolution, resolution test phantoms are used. The test phantoms are usually line pair phantoms or star patterns (Fig. 10-22) for projection radiography, and line pair phantoms can be purchased for CT as well.

The test phantoms are imaged, and the images are viewed to garner a qualitative understanding of the limiting spatial resolution of the imaging system. These routine measurements are often made on fluoroscopic equipment and for screen-film and digital radiographic systems.

10.3 NOISE

Figure 10-23 shows three isometric “images”; each one has similar contrast, but the amount of noise increases toward the right of the figure. Noise interjects a random

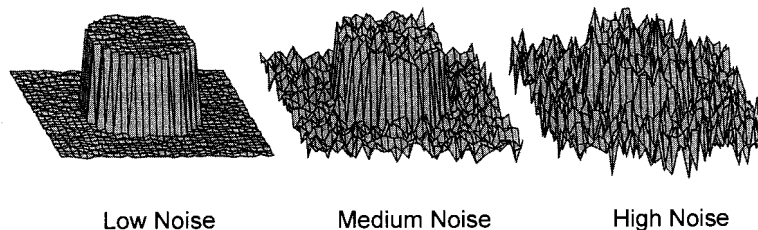


FIGURE 10-23. The concept of noise is illustrated using isometric display. A low noise image is shown on the left, with increasing amounts of noise added to the “images” toward the right.

or stochastic component into the image (or other measurement), and there are several different sources of noise in an image. There are different concepts pertaining to noise that are useful as background to this topic. Noise adds or subtracts to a measurement value such that the recorded measurement differs from the actual value. Most experimental systems have some amount of noise, and some have a great deal of noise. If we measure some value repeatedly, for instance the number of counts emitted from a weak radioactive source, we can compute the average count rate (decays/second), and there are ways to calculate the *standard deviation* as well. In this example, multiple measurements are made to try to improve the accuracy of the mean value because there are large fluctuations in each individual measurement. These fluctuations about the mean are stochastic in nature, and the determination of the standard deviation (σ) relates to the *precision* of the measurement.

As another example, suppose we go to a school district and measure the height of 1,000 children in the third grade. The individual measurements, if performed carefully, will have good precision. That is, if you measured the same child repeatedly, you'd get the same value with little or no deviation. Let's say we measure all 1,000 children, calculate the mean height of all the third graders, and we also compute the standard deviation. In this example, the standard deviation (σ) is not a measure of precision in the measurement, but rather is a measure of the natural variability in the height of children. So the standard deviation is a measure of the variability in some parameter, and that variability can be a result of random fluctuations in the measurement, or natural occurring fluctuations in the parameter itself.

Height measurements of a population of students can be represented by several parameters, the number of students measured, N , the mean height of the children, $\bar{X}$, and the standard deviation, σ . If we started with 1,000 individual height measurements of X_i , where X is the height (in cm) and the subscript i refers to each child ($i = 1$ to 1,000), then the mean is calculated as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N X_i \quad [10-14]$$

The standard deviation is computed as

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (\bar{X} - X_i)^2}{N - 1}} \quad [10-15]$$

The bar chart in Fig. 10-24 (marked h_i) is a *histogram* of the height measurement data for the $N = 1,000$ children. The x-axis is the height, and the number of children measured at each height is plotted on the y-axis. Using Equations 10-14 and 10-15, from the 1,000 measurements of X_i , we calculate $\bar{X} = 133.5$ cm and $\sigma = 11.0$ cm. The area under the curve is equal to $N = 1,000$.

Many types of data, including the heights of children, have a *normal* distribution. The normal distribution is also known as the *gaussian* distribution, and its mathematical expression is

$$G(x) = k e^{-\frac{1}{2} \left(\frac{\bar{X} - X}{\sigma} \right)^2} \quad [10-16]$$

The smooth curve (marked $G(x)$) shown in Fig. 10-24 illustrates a plot of the gaussian distribution, with $\bar{X} = 133.5$ cm and $\sigma = 11.0$ cm. Notice that these two para-

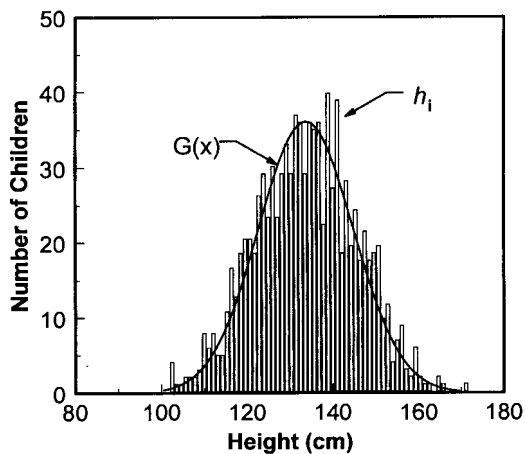


FIGURE 10-24. The bar chart (h_i) is a histogram of the height of 1,000 third grade children. The mean height is 133.5 cm and the standard deviation (σ) is 11.0 cm. A gaussian distribution [$G(x)$] is also illustrated with the same mean and standard deviation as the bar chart. For the gaussian curve, the mean corresponds to the center of the bell-shaped curve and the standard deviation is related to the width of the bell-shaped curve.

eters, $\bar{X}$ and σ , describe the shape of the gaussian distribution. The parameter x in Equation 10-16 is called the *dependent variable*, and x is the value corresponding to the x -axis in Fig. 10-24. The value of k is just a constant, used to normalize the height of the curve. The actual measured data, h_i , in Fig. 10-24, matches the shape of the gaussian distribution, $G(x)$, pretty well. If we continued to make more measurements (increasing N), the variations in the histogram h_i would decrease, and eventually it would smoothly overlay $G(x)$.

For the gaussian distribution, $\bar{X}$ describes where the center of the bell-shaped curve is positioned along the x -axis, and σ governs how wide the bell curve is. Figure 10-25 illustrates two gaussian curves, with means $\bar{X}_a$ and $\bar{X}_b$, and standard deviations σ_a and σ_b . The parameters $\bar{X}_a$ and σ_a correspond to third graders, and parameters $\bar{X}_b$ and σ_b correspond to sixth graders in this example. The point is that both parameters, the mean and the standard deviation, are useful in describing the shape of different measured data. In Fig. 10-25, the two gaussian curves describe (re-normalized) *probability density functions* (PDFs), in this case conveying the range of heights for the two different grades. For example, the left PDF in Fig. 10-25

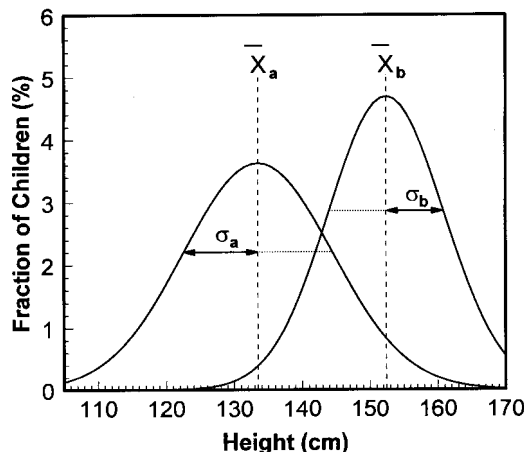


FIGURE 10-25. Two gaussians are illustrated with their corresponding means ($\bar{X}_a$ and $\bar{X}_b$) and standard deviations (σ_a and σ_b). These two curves illustrate the distribution of heights in third grade (curve A) and the sixth grade (curve B) in a certain school district. The point of this figure is to show that the gaussian distribution, with its two parameters $\bar{X}$ and σ , is quite flexible in describing various *statistical distributions*. Furthermore, the mean of a given distribution ($\bar{X}$) is not necessarily related or mathematically tied to the corresponding standard deviation (σ) of that distribution.

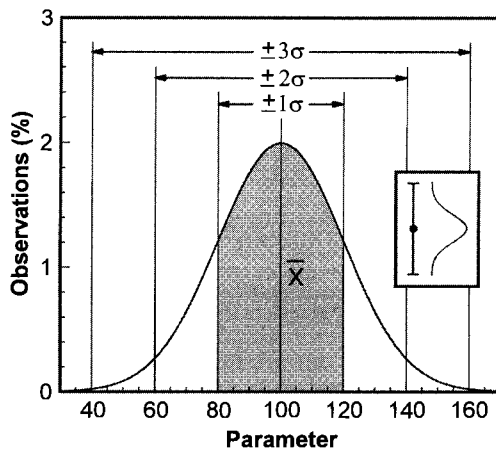


FIGURE 10-26. A gaussian distribution is illustrated with $\bar{X} = 100$ and $\sigma = 20$. This plot demonstrates a *probability density function*. The total area under the bell-shaped curve corresponds to 100% of the observations. What fraction of the observations would occur between $\bar{X} - \sigma$ and $\bar{X} + \sigma$? This probability is equal to the *area shaded gray* in the figure, and is 68%. The area corresponding to $\pm 2\sigma$ from the mean includes 95% of the observations, and $\pm 3\sigma$ from the mean includes 99% of all observations. Many scientific graphs make use of error bars, where the height of the error bar corresponds to $\pm 2\sigma$ around the mean (inset), denoting the *95% confidence limits*.

describes the probability that a third grader will have a certain height, if that third grader were selected randomly from the school district.

Figure 10-26 illustrates a generic gaussian PDF, with $\bar{X} = 100$ and $\sigma = 20$, and let's switch from the children's height example to a general description, and thus let the x-axis just be some parameter, and the y-axis be the number of observations made at each parameter value. Based on the PDF shown in Fig. 10-26, what is the probability that an observation will be within some range, say from $\bar{X} - \sigma$ to $\bar{X} + \sigma$? This probability is found by integrating the curve in Fig. 10-26 from $\bar{X} - \sigma$ to $\bar{X} + \sigma$, but conceptually this just corresponds to the area of the shaded region of the curve. For a gaussian-distributed parameter, where the total area of the curve is 100%, the area of the curve (and hence the total probability of observations) between $\pm 1\sigma$ is 68%, the area between $\pm 2\sigma$ is 95%, and 99% of the observations occur between $\pm 3\sigma$. These values are marked on Fig. 10-26. Most scientific publications make use of plots with error bars (inset on Fig. 10-26), and often the error bars are plotted to include $\pm 2\sigma$, thereby spanning the *95% confidence interval*.

It is worth mentioning that although the gaussian distribution is a very common form of a probability density function, it is not the only form. Figure 10-27A illustrates a uniformly distributed PDF, and this corresponds to, for example, the probability of an x-ray photon hitting a certain position (x) along a pixel in a digital detector. Figure 10-27B illustrates the angular distribution of a 20-keV x-ray photon being scattering by Compton interaction. Another common form of PDF is defined by the Poisson distribution, $P(x)$, which is expressed as:

$$P(x) = \frac{m^x}{x!} e^{-m} \quad [10-17]$$

where m is the mean and x is the dependent variable. The interesting thing about the Poisson distribution is that its shape is governed by only one variable (m), not two variables as we saw for the gaussian distribution. Figure 10-28 illustrates the Poisson distribution with $m = 50$ (dashed line), along with the gaussian distribution (Equation 10-16) with $\bar{X} = 50$ and $\sigma = \sqrt{\bar{X}}$. In general, the Poisson distribution is exceedingly similar (not identical) to the gaussian distribution when σ is set such that $\sigma = \sqrt{\bar{X}}$. Indeed, the gaussian equation is often used to approximate the Poisson distribution with the above stipulation, because the factorial ($x!$, i.e., $5! = 1 \times 2$

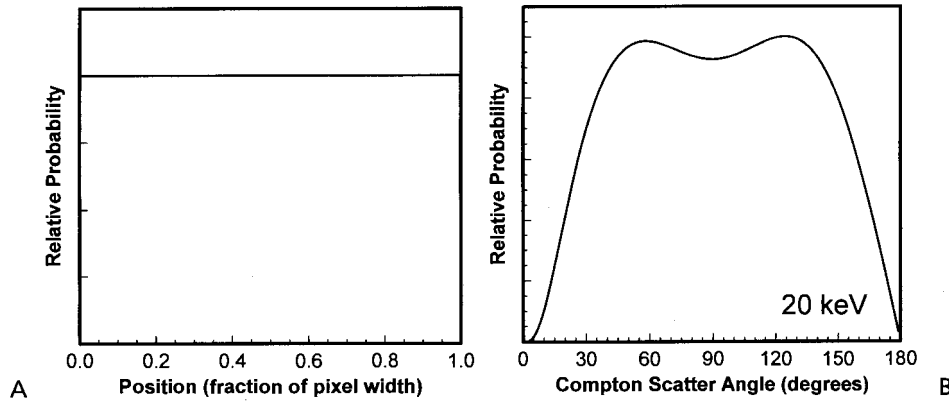


FIGURE 10-27. Not all probability density functions (PDFs) are distributed according to the gaussian distribution. **(A)** The probability of an x-ray photon landing at some position along the width of a pixel demonstrates a uniformly distributed PDF. **(B)** The probability that a 20-keV x-ray photon will be scattered at a given angle by the Compton interaction is also illustrated. In nature, and especially in physics, there are many shapes of PDFs other than the gaussian. Nevertheless, the gaussian distribution is the most commonly applicable PDF, and many measured phenomena exhibit gaussian characteristics.

$\times 3 \times 4 \times 5$) in Equation 10-17 makes computation difficult for larger values of x (for example, $69! = 1.7 \times 10^{98}$).

X-ray and γ -ray counting statistics obey the Poisson distribution, and this is quite fortunate. Why? We saw that the gaussian distribution has two parameters, X , and σ , which govern its shape. Knowing one parameter does not imply that you know the other, because there is no fixed relationship between σ and X (as the children's height example was meant to illustrate). However, with the Poisson distribution (actually its gaussian approximation), knowing the mean (X) implies that σ is known as well, since $\sigma = \sqrt{X}$. In other words, σ can be predicted from X . Why is this a fortunate situation for x-ray and γ -ray imaging? It is fortunate because it means that we can adjust the noise (σ) in an image by adjusting the (mean) number of photons used to produce the image. This is discussed more in the next section.

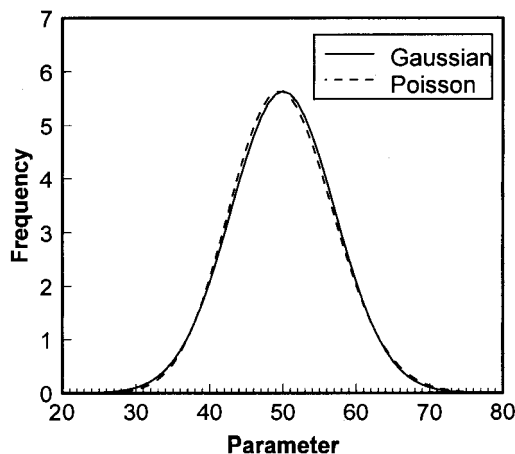


FIGURE 10-28. The Poisson distribution has but one parameter, which describes its shape, m , unlike the gaussian distribution, which is described by two parameters (X and σ). However, the gaussian distribution represents an excellent approximation to the Poisson distribution when $\sigma = \sqrt{X}$.

Quantum Noise

The term *quantum* is defined as “something that may be counted or measured,” and in medical imaging *quanta* is a generic term for signals that come in individual, discrete packets. Examples in medical imaging of quanta are x-rays, electrons, ions, and light photons. Other examples could include bricks, marbles, grains of sand, and anything that can be counted in a discrete manner. When discussing the *number* of x-ray quanta, it is common nomenclature to use the symbol N as the mean number of photons per unit area (whereas the symbol X was used in the previous section). For a digital x-ray detector system with square pixels (as a simple example), if the average number of x-rays recorded in each pixel is N , then the noise (per pixel) will be

$$\sigma = \sqrt{N} \quad [10-18A]$$

Or stated differently,

$$\sigma^2 = N \quad [10-18B]$$

where σ is called the *standard deviation* or the *noise*, and σ^2 is the *variance*. Noise as perceived by the human observer in an image is the *relative noise*, also called the coefficient of variation (COV):

$$\text{Relative noise} = \text{COV} = \frac{\sigma}{N} \quad [10-19]$$

Table 10-1 illustrates how the relative noise changes as N (photons/pixel) is increased. As N increases, σ also increases but more slowly (by the square root). Consequently, the *relative noise*, σ/N , *decreases with increasing N* . The inverse of the relative noise is the signal-to-noise ratio (SNR), N/σ . Notice that

$$\text{SNR} = \frac{N}{\sigma} = \frac{N}{\sqrt{N}} = \sqrt{N} \quad [10-20]$$

Thus as N is increased in an image, the SNR increases as $\sqrt{N}$. When the number of x-ray quanta N is increased, the radiation dose to the patient also increases. If N is doubled (at the same effective x-ray energy), then the radiation dose to the patient will also double. Image quality, however, is largely determined by the SNR. When N is doubled, the SNR increases by a factor of $\sqrt{2}$. So, to double the SNR, the dose to the patient needs to be increased by a factor of 4. To reiterate, these predictions can be made with confidence only because we know the Poisson distribution is at play when dealing with photon counting statistics, and this allows σ to be predicted from N .

TABLE 10-1. EXAMPLES OF NOISE VERSUS PHOTONS

Photons/Pixel (N)	Noise (σ) ($\sigma = \sqrt{N}$)	Relative Noise (σ/N) (%)	SNR (N/σ)
10	3.2	32	3.2
100	10	10	10
1,000	31.6	3.2	32
10,000	100	1.0	100
100,000	316.2	0.3	316

SNR, signal-to-noise ratio.

In all medical x-ray imaging, we want to reduce the radiation dose to the patient to a minimum level, and we saw that there is a trade-off between SNR and radiation dose. Because of this concern, it is very important that every x-ray photon that reaches the detector be counted and contribute toward the SNR. Perfect x-ray detector systems are ones that are “x-ray quantum limited,” that is, where the image SNR is dictated by only the SNR of the x-ray quanta used. While this is the goal, real-world detectors are not perfect and they generally do not absorb (detect) all the x-rays incident upon them. For example, a screen-film detector system may detect (depending on x-ray energy) about 60% of the x-ray photons incident on it. The quantum detection efficiency (QDE) is the ratio of the number of detected photons (photons absorbed by the detector) to the number of incident photons for a given detector system:

$$QDE = \frac{N_{\text{absorbed}}}{N_{\text{incident}}} \quad [10-21]$$

In terms of the image SNR, it is the number of *detected* photons, not incident photons, that are used in the SNR calculation, such that

$$SNR_{\text{reality}} = \sqrt{N_{\text{detected}}} = \sqrt{QDE \times N_{\text{incident}}} \quad [10-22]$$

Equation 10-22 includes the influence of the DQE, but does not include the influence of other sources of noise in the detector system. This will be discussed in the next section.

Other Sources of Image Noise

An x-ray imaging system employing an intensifying screen and a CCD camera is illustrated in Fig. 10-29A. As a single x-ray photon stimulates the imaging system, various numbers of quanta are used at different stages in the system. The stages are described in Table 10-2, and the number of quanta propagating through two dif-

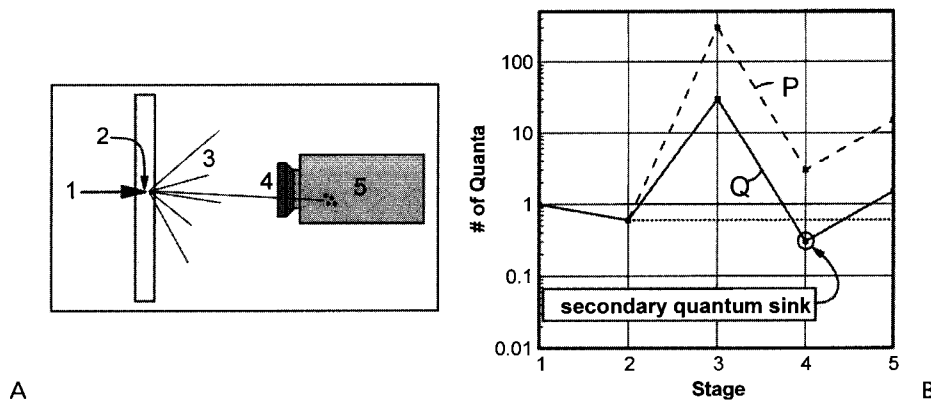


FIGURE 10-29. A: An imaging system is illustrated with different stages, 1 through 5 (see Table 10-2 for a description of the different stages). **B:** The *quantum accounting diagram* is illustrated for two different imaging systems, P and Q. In the quantum accounting diagram, if at any point the number of quanta is lower than the number of detected x-ray photons, a *secondary quantum sink* occurs. System Q has a lower gain at stage 3 than system P, and a secondary quantum sink occurs at stage 4 as a result.

TABLE 10-2. STAGES OF SIGNAL TRANSFER (SEE FIG. 10-29)

Stage	Description	System P	System Q
1	Incident x-ray photon	1	1
2	Detected x-ray photons	0.6	0.6
3	Light photons emitted by screen	30	300
4	Photons entering camera lens	0.3	3
5	Electrons generated in CCD	1.5	15

ferent systems (P and Q) are tabulated. The number of quanta at each stage for both systems is plotted in Fig. 10-29B, which is called a *quantum accounting diagram*. For an x-ray imaging system to be *x-ray quantum limited*, a desirable goal, the number of quanta used along the imaging chain cannot be less than the number of detected x-ray quanta. System P has excellent light conversion efficiency (500 light photons per detected x-ray photon), and the system behaves in an x-ray quantum limited manner, because at no point along the imaging chain does the number of quanta dip below the number of detected x-rays. System Q, however, has a poor light conversion efficiency (50), and at stage 4 (number of light photons entering lens of the CCD camera) the number of quanta dips below the number of detected x-ray quanta (dotted line). System Q has a *secondary quantum sink*, and although the number of quanta is amplified in the CCD, the system will nevertheless be fundamentally impaired by the quantum sink at stage 4. The reason why a secondary quantum sink is to be avoided is that it means that the noise properties of the image are not governed by the x-ray quanta, as they should be, but by some design flaw in the system. Such a system wastes x-ray photons, and thus causes unnecessary radiation dose to the patient. If the CCD camera in Fig. 10-29A was replaced with the radiologist's eye, the system would be exactly that used in old-fashioned fluorography, as practiced in the 1930s and 1940s, a defunct imaging system well known to be very dose inefficient.

Contrast Resolution

The ability to detect a low-contrast object on an image is highly related to how much noise (quantum noise and otherwise) there is in the image. As noise levels decrease, the contrast of the object becomes more perceptible. The ability to visualize low-contrast objects is the essence of *contrast resolution*. Better contrast resolution implies that more subtle objects can be routinely seen on the image. Contrast resolution is very much related to the SNR, which was discussed at length above. A scientist working on early television systems, Albert Rose, showed that to reliably identify an object, the SNR needed to be better than about 5. This requirement has become known as *Rose's criterion*. It is possible to identify objects with SNRs lower than 5.0; however, the probability of detection is lower than 100%. Figure 10-30 (top row) illustrates a low-contrast circle on a homogeneous noisy background, with the amount of noise increasing toward the right of the figure. Figure 10-30A has higher SNR than Fig. 10-30C, and it demonstrates better contrast resolution. In this example, the contrast of the circle was increased, which increased the SNR. The SNR will also increase if the area of the circle is increased, with no change in the contrast. The lower row of images on Fig. 10-30

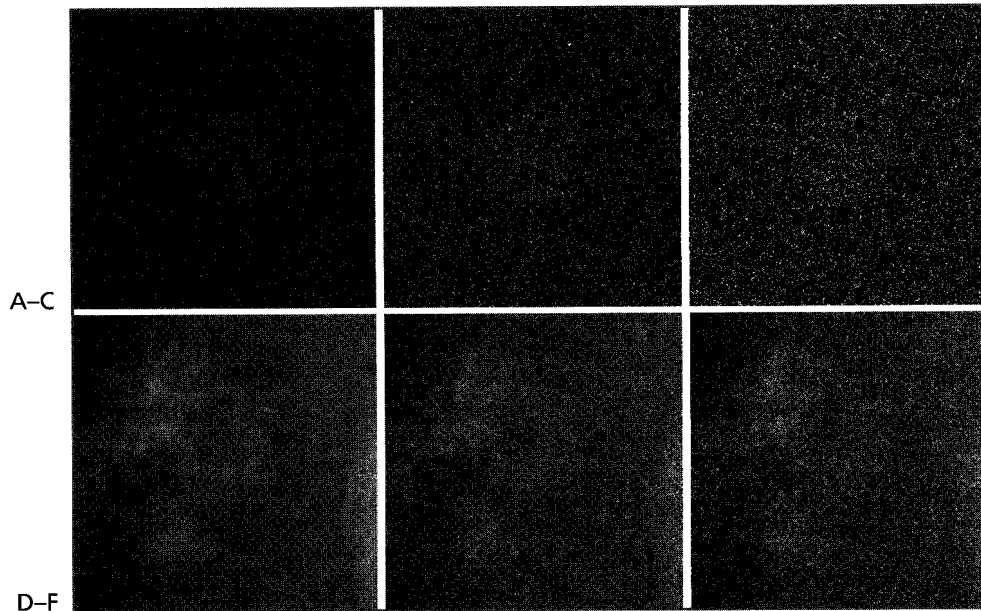


FIGURE 10-30. A circular signal is shown on a constant noisy background (**A–C**), with the signal to noise ratio (SNR) decreasing toward panel **C**. The lower panels (**D–F**) show the same circle and noise, with a breast parenchyma pattern overlaid. The presence of normal anatomic contrast acts as a source of structure noise, and reduces the detectability of the circle in this example.

(**D–F**) is similar to the upper row, except that normal anatomic tissue structure from a mammogram has been added to the background. Normal tissue anatomy can act to mask subtle lesions and reduces contrast resolution, and this effect is called *structure noise*. The added structure noise from the normal anatomy reduces the ability to visualize the circle.

Noise Frequency: $W(f)$

If an image is acquired with nothing in the beam, a contrast-less, almost uniformly gray image is obtained. The gray isn't exactly uniform, due to the presence of noise. This type of image is used to calculate noise, for instance the standard deviation σ is calculated from a uniform image using Equation 10-15. To the human eye, often the noise appears to be totally random; however, there usually are subtle relationships between the noise at one point and the noise at other points in the image, and these relationships can be teased out using noise frequency analysis. An example is given below to illustrate how random noise can have nonrandom components to it.

Imagine that you are standing near a tall glass wall with the ocean on the other side, and you are watching the water level (water height on the wall along a particular vertical line) fluctuate as the waves splash against the wall. At first glance, the water seems to splash around in a random way, but we then set out to analyze this by measuring the height of the water against the wall continuously for a 24-hour period. These measurements are plotted in Fig. 10-31A. The sinusoidal background

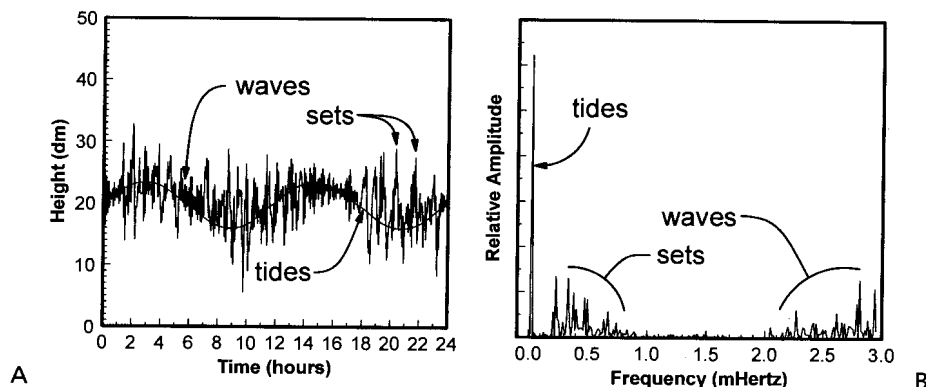


FIGURE 10-31. A: The height (1 dm = 10 cm) of the ocean against a glass wall is monitored for a 24-hour period. Large sinusoidal fluctuations corresponding to tidal changes are seen, and overlaid on the tidal background are large amplitude infrequent fluctuations (wave sets) and smaller amplitude more frequent ones (waves). **B:** Fourier analysis of the trace in **A** reveals the frequency of the various surf components.

of tides is seen, as are spikes representing the occasional big sets of waves rolling in, and the ordinary ocean waves represent the short-duration fluctuations on this plot. Figure 10-31B illustrates the results when the Fourier transform is computed on these data. In this plot, the different frequencies of the various wave and tidal phenomena can be seen. The tides are very low frequency but are high in amplitude (tidal variations are usually larger than wave height). The tides experience about two complete cycles per 24-hour period, corresponding to a frequency of 1 cycle/12 hours = 2.3×10^{-5} hertz (very close to zero on this plot). The wave sets show up in the general vicinity of 0.0002 to 0.0005 hertz, corresponding to a periodicity of around 50 minutes. The waves occur much more often, corresponding to their higher frequency.

Noise on an image often has frequency components to it as well, and a frequency analysis similar to the ocean wave example is performed. Just like the Fourier transform of how the imaging system passes "signal" (i.e., the LSF) leads to the calculation of the MTF, the Fourier transform of how an imaging system passes noise leads

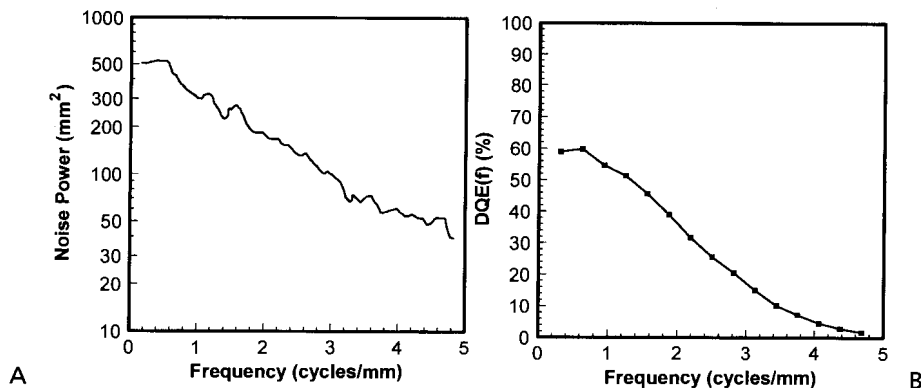


FIGURE 10-32. A: A noise power spectrum. **B:** A plot of the detective quantum efficiency (DQE).

to the calculation of the noise power spectrum [NPS(f)]. (The mathematical details are beyond the scope of this text.) The NPS(f) (also called the Wiener spectrum) is the noise variance (σ^2) of the image, expressed as a function of spatial frequency (f). An NPS(f) is shown for an x-ray detector system in Fig. 10-32A.

10.4 DETECTIVE QUANTUM EFFICIENCY

The detective quantum efficiency (DQE) of an imaging system is a very useful description of an x-ray imaging system, used by imaging scientists, to describe the overall SNR performance of the system. In conceptual terms, the DQE can be defined as the ratio of the SNR^2 output from the system, to the SNR^2 of the signal input into the system:

$$\text{DQE} = \frac{\text{SNR}_{\text{OUT}}^2}{\text{SNR}_{\text{IN}}^2} \quad [10-23]$$

The MTF(f) describes how well an imaging system *processes signal*, and the NPS(f) describes how well an imaging system *processes noise* in the image. Combining these concepts, the numerator of Equation 10-23 is given by

$$\text{SNR}_{\text{OUT}}^2 = \frac{[\text{MTF}(f)]^2}{\text{NPS}(f)} \quad [10-24]$$

The NPS(f) is the noise variance [i.e., $\sigma^2(f)$], so it is already squared. The SNR_{IN} to an x-ray imaging system is simply $\text{SNR}_{\text{IN}} = \sqrt{N}$ (Equation 10-20), and thus $\text{SNR}_{\text{IN}}^2 = N$, and Equation 10-23 becomes

$$\text{DQE}(f) = \frac{k [\text{MTF}(f)]^2}{N \text{NPS}(f)} \quad [10-25]$$

where k is a constant that converts units, and its value will not be discussed here. The NPS(f) is determined from an image (or images) acquired at a mean photon fluence of N photons/mm². The DQE(f) has become the standard by which the performance of x-ray imaging systems is measured in the research environment. With digital radiographic and fluoroscopic x-ray detectors becoming ever more common, the DQE(f) is likely to emerge as a quality assurance metric over time. Note that the DQE(f) is an excellent description of the dose efficiency of an x-ray detector system. The detective quantum efficiency, DQE(f), should not be confused with the quantum detection efficiency, the QDE. There is a relationship between the two, however. For a system with ideal noise properties, the QDE(f) will be equal to the DQE (at $f = 0$). The DQE(f) of an x-ray detector system is shown in Fig. 10-32B.

10.5 SAMPLING AND ALIASING IN DIGITAL IMAGES

An array of detector elements (Fig. 10-33) has two distances that are important, the *sampling pitch* (or pixel pitch) and the *detector aperture width*. When a signal strikes a digital detector system (for example x-rays landing on an x-ray detector, ultrasound echoing back to the transducer, etc.), each individual detector records only the signal incident upon it. For the many detector systems using square, contiguous

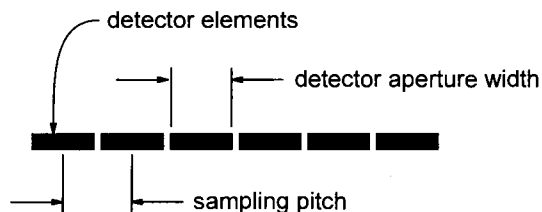


FIGURE 10-33. Detector aperture width and sampling pitch.

detector elements ("dels"), the sampling distance between dels is virtually identical to the aperture width (from edge to edge) of the dels, and this can cause confusion. Two very different phenomena occur when digital detector systems are used, or when an analog film is digitized: (a) The analog image data (signal) is *sampled*, and the sampling distance is equal to the center-to-center spacing *between* the individual detector elements. (2) Due to the finite width of the detector element, *signal averaging* will occur over the *detector aperture width*. Many detector systems are two-dimensional, but for convenience these two effects will be described referring only to one dimension.

Sampling Requirements and Aliasing

The spacing between samples on a digital image determines the highest frequency that can be imaged. This limitation is called the *Nyquist criterion* or Nyquist limit. If the spacing between samples (*pixel pitch* or *sampling pitch*) is Δ , then the Nyquist frequency (F_N) is given by:

$$F_N = \frac{1}{2\Delta} \quad [10-26]$$

The Nyquist criterion essentially states that our conceptual sinusoidal input signal needs to be sampled such that each cycle is sampled at least twice. If a frequency component in an image exceeds the Nyquist frequency, it will be sampled

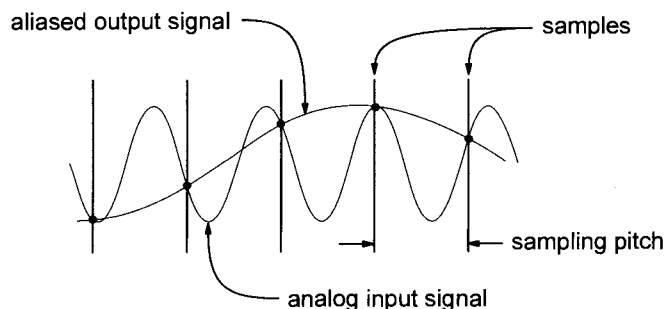


FIGURE 10-34. The geometric basis of aliasing. An analog input signal of a certain frequency F is sampled with a periodicity shown by the vertical lines ("samples"). The interval between samples is wider than that required by the Nyquist criterion, and thus the input signal is *undersampled*. The sampled points (solid circles) are consistent with a much lower frequency signal ("aliased output signal"), illustrating how undersampling causes high input frequencies to be recorded as aliased, lower measured frequencies.

less than twice per cycle, and it will be *aliased*. When a high-frequency signal becomes aliased, it wraps back onto the image as a lower frequency. Figure 10-34 illustrates graphically why and how aliasing occurs. Aliasing on an image often looks like a moiré pattern. Aliasing can occur in space or in time. For example, if an angiographic sequence is acquired of a pediatric heart beating faster than the image acquisition rate, it will appear during a dynamic display of the imaging sequence to be beating much slower than in reality, a case of temporal aliasing. The wagon wheels on TV westerns seemingly rotate more slowly and sometimes in the wrong direction (a phase effect) as a result of temporal aliasing.

We learned earlier that the Fourier transform is a mathematical technique for breaking a signal down into a sine wave representation of an image. If a pure sine wave of, for example, 3 cycles/mm is input into an imaging system and then the Fourier transform is computed on that (one-dimensional) image, there will be a sharp peak at 3 cycles/mm. This is shown in Fig. 10-35. The three graphs on the left side of Fig. 10-35 illustrate frequency diagrams for input frequencies (F_{in}) of 2, 3, and 4 cycles/mm. The Nyquist limit on the imaging system modeled in Fig. 10-35 was 5 cycles/mm, because the detector spacing was 0.10 mm (Equation 10-26). In the three graphs on the right side of Fig. 10-35, the input frequencies exceed the Nyquist limit, as indicated. Computing the Fourier transform on these images reveals sharp peaks at lower frequencies than those input; in fact, these aliased frequencies *wrap around* the Nyquist frequency. For example, when the input frequency is 6 cycles/mm, which is 1 cycle/mm *higher* than the Nyquist frequency, the recorded aliased frequency is 4 cycles/mm, 1 cycle/mm *lower* than Nyquist. Similarly, for an input of 7 ($5 + 2$) cycles/mm, the aliased frequency becomes 3 ($5 - 2$) cycles/mm.

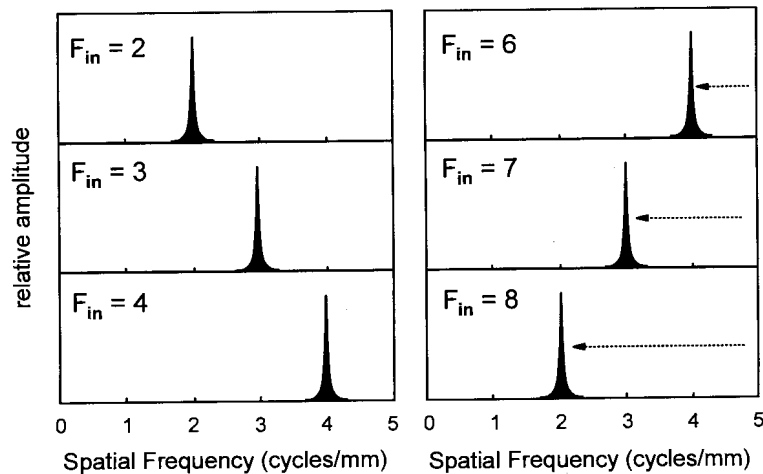


FIGURE 10-35. Each of the six panels illustrates the Fourier transform of a sine wave with input frequency F_{in} , after it has been detected by an imaging system with a Nyquist frequency, F_N , of 5 cycles/mm (i.e., sampling pitch = 0.10 mm). **Left panels:** Input frequencies that obey the Nyquist criterion ($F_{in} < F_N$), and the Fourier transform shows the recorded frequencies to be equal to the input frequencies. **Right panels:** Input frequencies that do not obey the Nyquist criterion ($F_{in} > F_N$), and the Fourier transform of the signal indicates recorded frequencies quite different from the input frequencies, demonstrating the effect of *aliasing*.

Since most patient anatomy does not have a high degree of periodicity, rarely does aliasing become a noticeable problem in x-ray imaging in terms of structure in the patient. However, antiscatter grids have a high degree of periodicity, and grids will cause severe moiré effects on digital radiographic or fluoroscopic images if positioned incorrectly, or if the grid strip density is too low. In MRI, aliasing can occur because the anatomic signal is highly periodic [the radiofrequency (RF) is sinusoidal], and if undersampling occurs, artifacts will result. Ultrasound imaging makes use of sinusoidal acoustic waves, and aliasing can also be realized in clinical images, especially in Doppler applications.

Aperture Blurring

The second consequence of recording an analog signal with a digital detector system is *signal averaging* across the detector aperture. Figure 10-36A illustrates an analog signal, and each detector sums the signal across its width and records this averaged value. The aperture width determines the distance in which the signal is averaged, and seen in a one-dimensional profile (i.e., a line spread function), the aperture has the appearance of a rectangle (Fig. 10-36). The rectangle-shaped LSF is called a *rect* function, and it causes imaging blurring for exactly the same reasons that a more traditional bell-shaped LSF does. As discussed previously, the MTF of the imaging system is the Fourier transform of the LSF, and computing the FT of a rect function results in what is called a *sinc* function, where

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \quad [10-27]$$

The rectangular-shaped aperture of the detector elements of width a , causes a blurring that is described on the MTF by the *sinc* function, such that

$$\text{MTF}(f) = \frac{\sin(\pi f a)}{\pi f a} \quad [10-28]$$

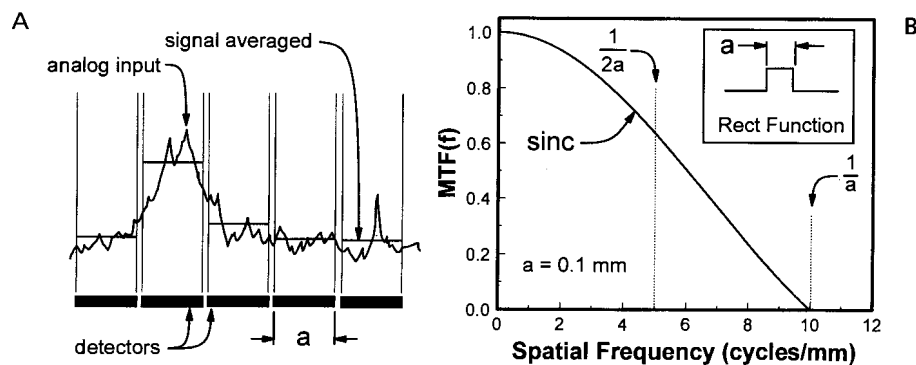


FIGURE 10-36. Most detectors have a finite aperture width, a . The signal that is incident upon the face of each detector element is averaged over that aperture. This averaging is mathematically equivalent to convolving the input signal with a rectangular spread function (a *rect* function) of width a . In the MTF, this averaging is described by the *sinc* function, which is the Fourier transform of the *rect*. For a *rect* function of width a , the MTF will become zero at $F = 1/a$. However, if the sampling pitch is also equal to a (Fig. 10-33), the Nyquist limit will be $1/(2a)$, and aliasing will occur for frequencies between $1/(2a)$ and $1/a$.

The *sinc*-shaped MTF for a detector aperture of $a = 0.100$ mm is shown in Fig. 10-36B.

10.6 CONTRAST-DETAIL CURVES

As discussed in previous sections of this chapter, the spatial resolution is best described by the modulation transfer function, and the contrast resolution is best described by the SNR or by the noise power spectrum of the imaging system. The detective quantum efficiency [DQE(f)] is a good *quantitative* way to combine concepts of the spatial and contrast resolution properties of an imaging system. An excellent *qualitative* way in which to combine the notions of spatial and contrast resolution is the *contrast-detail curve*. A typical contrast-detail curve is shown in Fig. 10-37A, where the image was acquired from a *C-D phantom* as shown. The x-axis of the image corresponds to the size of objects (*detail*), with smaller objects toward the left, and the y-axis corresponds to the contrast of the objects, with lower contrast toward the bottom of the image. As objects get smaller and lower in contrast, their SNR is reduced and they become harder to see on the image. The white line on the image (Fig. 10-37A) corresponds to the transition zone, where objects above and to the right are visualized, and objects below and to the left of the line are not seen.

The utility of C-D diagrams is illustrated in Fig. 10-37B, where the contrast-detail curves for two systems (A and B) are shown. System A has better spatial resolution than system B, because the curve extends further to the left toward smaller detail (vertical dashed lines compare the locations of systems A and B on the detail axis). Similarly, system B has better contrast resolution. The curves on C-D diagrams indicate the transition from objects that can be seen to those that can't, and these curves are derived subjectively by simple inspection. It is reasonable to expect some interobserver variability in the development of contrast-detail curves, but as a quick subjective evaluation of an imaging system, they are quite useful. Contrast-

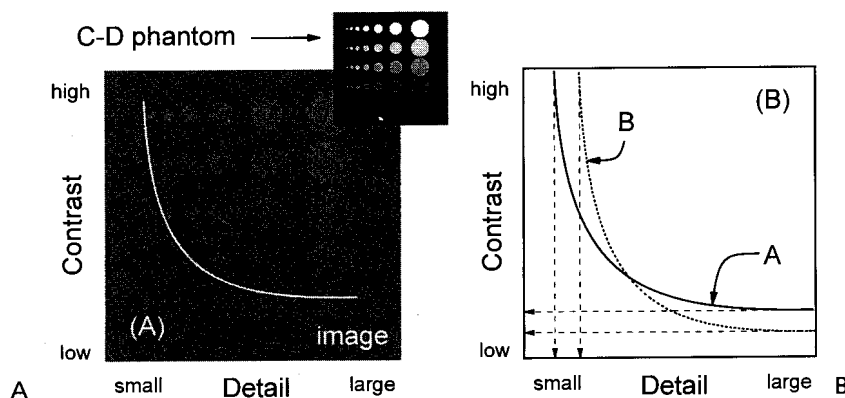


FIGURE 10-37. A: A contrast-detail phantom and an image of the C-D phantom. Objects that are small and have low contrast will be hard to see (*lower left* of C-D image), and objects that are large and have high contrast will be easy to see (*upper right*). The *white line* is the demarcation line between the circular test objects that can and can't be seen. Contrast-detail curves are frequently just presented as plots **(B)**. **B:** System A has higher spatial resolution but lower contrast resolution, compared to system B.

detail curves are commonly used in CT, but the concept applies to MRI, ultrasound, radiography, and other types of imaging systems as well.

10.7 RECEIVER OPERATING CHARACTERISTIC CURVES

Receiver operating characteristic (ROC) curve analysis was developed many years ago to assess the performance of radar detectors, and the same techniques have been widely adapted to radiology images. ROC analysis can be performed on human observers, and this analysis tool is useful in comparing the diagnostic abilities of different radiologists (e.g., comparing a resident to a faculty radiologist), or it can be used to compare the diagnostic performance of two different imaging modalities (e.g., MRI versus CT).

In any diagnostic test, the most basic task of the diagnostician is to separate abnormal subjects from normal subjects. The problem is that in many cases, there is significant overlap in terms of the appearance on the image. For example, some abnormal patients have normal-looking films, and some normal patients have abnormal-looking films. Figure 10-38A illustrates the situation, where there is a normal and abnormal population, and overlap occurs. The two bell-shaped curves are histograms, and they show the number of patients on the y-axis and the value of some decision criterion on the x-axis. The *decision criterion* is a vague term, and it really changes with the specific task. As one example, the decision criterion could be the CT number of a pulmonary nodule, as it is known that higher CT numbers correspond to a more calcified lesion, and are more likely to be benign. In this example, the decision criterion would be how *low* the CT number is, since normals (patients with benign disease) are shown on the left and abnormals (patients with lung cancer) on the right. In most radiology applications, the decision criterion is

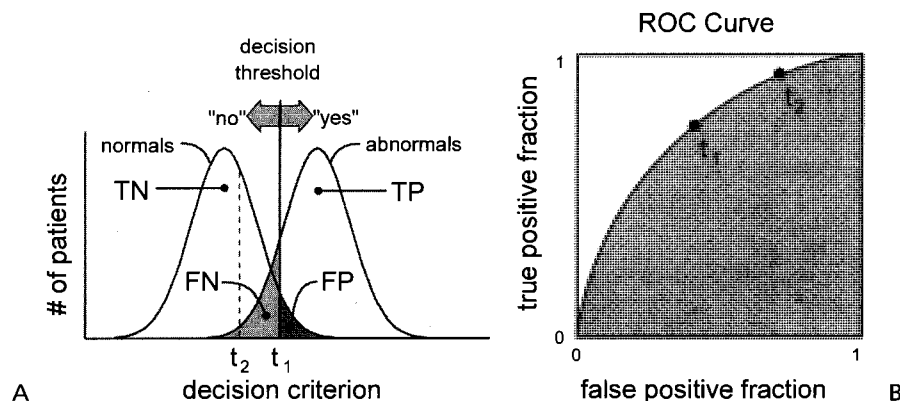


FIGURE 10-38. A: The basics of signal detection theory. The study population includes normals and abnormals, and they are histogrammed according to the *decision criterion* (see text). The diagnostician applies a decision threshold, and calls cases to the right of it abnormal ("yes") and to the left of it negative ("no"). The true positive, true negative, false positive, and false negatives can then be computed. These values are used to calculate the true-positive fraction and the false-positive fraction, which are plotted on the ROC curve (**B**). The decision threshold at each point (t_1) gives rise to one point on the ROC curve (t_1). Shifting the decision threshold toward the left (to the dashed line marked t_2 on the left figure) will increase the true-positive fraction but will also increase the false-positive fraction.

not just one feature or number, but is rather an overall impression derived from a number of factors, the *gestalt*.

While overlap occurs between normals and abnormals (Fig. 10-38A), the radiologist is often charged with diagnosing each case one way or another. The patient is either normal (“no”) or abnormal (“yes”), and from experience the diagnostician sets his or her own *decision threshold* (Fig. 10-38A) and based on this threshold, the patient is diagnosed as either normal or abnormal. The so-called 2×2 *decision matrix* follows from this:

THE 2×2 DECISION MATRIX

	Actually Abnormal	Actually Normal
Diagnosed as Abnormal	True Positive (TP)	False Positive (FP)
Diagnosed as Normal	False Negative (FN)	True Negative (TN)

The 2×2 decision matrix essentially defines the terms *true positive*, *true negative*, *false positive*, and *false negative*. For a single threshold value and the population being studied (the two bell curves in Fig. 10-38A), a single value for TP, TN, FP, and FN can be computed. The sum $TP + TN + FP + FN$ will be equal to the total number of normals and abnormals in the study population. Notice that to actually calculate these parameters, “truth” needs to be known. Most of the work in performing ROC studies is the independent confirmation of the “true” diagnosis, based on biopsy confirmation, long-term patient follow-up, etc.

From the values of TP, TN, FP, and FN, the true-positive fraction (TPF), also known as the *sensitivity*, can be calculated as

$$TPF = \frac{TP}{TP + FN} \quad [10-29]$$

The false-positive fraction (FPF) is calculated as

$$FPF = \frac{FP}{FP + TN} \quad [10-30]$$

A ROC curve is a plot of the true-positive fraction versus the false-positive fraction, as shown in Fig. 10-38B. A single threshold value will produce a single point on the ROC curve (e.g., t_1), and by changing the threshold value other points can be identified (e.g., t_2). To produce the entire ROC curve (a series of points), the threshold value must be swept continuously along the decision criterion axis of Fig. 10-38A. In practice with human observers, about five different points on the curve are realized by asking the human observer his or her level of confidence (*definitely there*, *maybe there*, *uncertain*, *maybe not there*, and *definitely not there*), and the rest of the ROC curve is interpolated. The different confidence levels correspond to five different decision threshold values.

Other important terms can be computed from the values of TP, TN, FP, and FN. The *sensitivity* is the fraction of abnormal cases that a diagnostician actually calls abnormal:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad [10-31]$$

The *specificity* is the fraction of normal cases that a diagnostician actually calls normal:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad [10-32]$$

Many medical tests are characterized only by their sensitivity and specificity. An ROC curve is a much richer description of the performance potential of a diagnostic test, because it describes diagnostic performance as a function of the decision threshold value. The ROC curve is a plot of sensitivity (TPF) versus (1 – specificity) (or FPF), and thus sliding along the curve allows one to trade off sensitivity for specificity, and vice versa. Why is this important? Say that two radiologists are equally skilled and have the same ROC curve for interpreting mammograms, for example. If one of the radiologists has had a bad experience with a lawsuit, she may subconsciously (or consciously) slide her decision threshold toward better sensitivity (by overcalling even mildly suspicious lesions), compromising specificity. Her benign biopsy rate would increase relative to that of her colleague, and her sensitivity and specificity would be different. She nevertheless performs with equal diagnostic ability.

Accuracy is defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad [10-33]$$

It is tempting to use *accuracy* as a measure of diagnostic performance, but accuracy is highly influenced by disease *incidence*. For example, a mammographer could invariably improve accuracy by just calling all cases negative, not even needing to look at the films. In such a scenario, because the incidence of breast cancer in the screening population is low (~0.003), with all cases called negative (normal), TN would be very large, FP and TP would be zero (since all cases are called negative, there would be no positives), and accuracy reduces to TN/(TN + FN). Since TN>>>FN, accuracy would approach 100%. Although the specificity would be 100%, the sensitivity would be 0%, so clearly accuracy is a poor metric of the performance of a diagnostic test.

The *positive predictive value* refers to the probability that the patient is actually abnormal, when the diagnostician *says* the patient is abnormal.

$$\text{Positive predictive value} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad [10-34]$$

Conversely, the negative predictive value refers to the probability that the patient is actually normal, when the diagnostician *says* the patient is normal.

$$\text{Negative predictive value} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad [10-35]$$

Interpretation

The mechanics of ROC curves were outlined above, but how are ROC curves used and what do they convey? An ROC curve is essentially a way of analyzing the SNR associated with a certain diagnostic task. In addition to the inherent SNR of the imaging modality under investigation, different human observers have what's called *internal noise*, which affects individual performance, and therefore different radiologists may have different ROC curves. Figure 10-39A shows three sets of histograms; set A shows a situation where there is almost complete overlap between

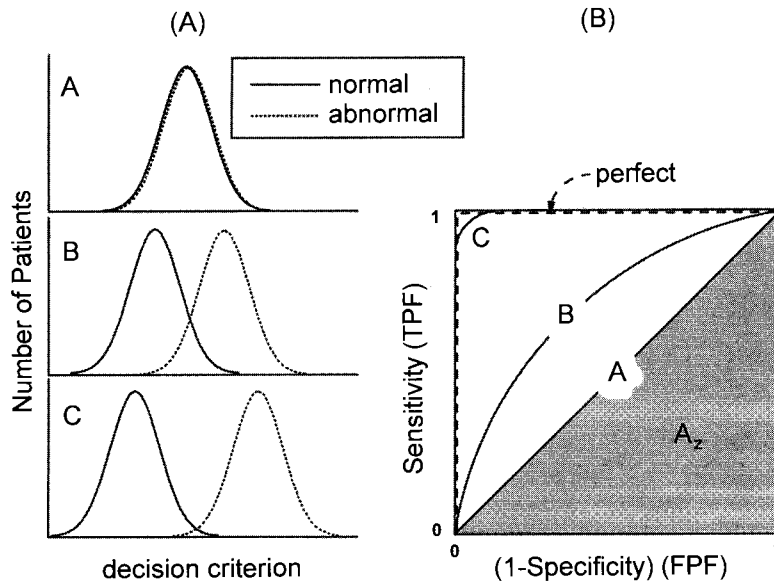


FIGURE 10-39. ROC curves are a way of expressing the signal-to-noise ratio (SNR) of the diagnostic test being studied. If the SNR of the test is close to zero, the normal and abnormal curves will overlay each other, and this would lead to an ROC curve marked A (in **B**). As the SNR increases, the ROC curve performance also improves (curves B and C in **B**). A perfect ROC curve rides the left and top edges of the ROC plot. Pure guessing is the *diagonal* marked as A. The area under the ROC curve, commonly referred to as A_z , is a good quantitative description of the detection performance of the diagnostic test.

abnormal and normal cases, and this leads to an ROC curve on Fig. 10-39B marked A. With case A, the SNR is near zero ($\sim 100\%$ overlap), and ROC curve A represents *pure guessing* in terms of the diagnosis. As the separation between normal and abnormal cases increases (with B and C on Fig. 10-39), the corresponding ROC curves approach the upper left corner. Set C has a good deal of separation between normal and abnormal, and ROC curve C approaches perfect performance. A perfect ROC curve is indicated as the dashed curve on Fig. 10-39B.

Both axes on the ROC curve scale from 0 to 1, and therefore the area of the box is 1. The diagonal associated with pure guessing (worst performance) bisects the box (Fig. 10-39B), and the area under it is 0.5. Best performance corresponds to an area of 1. The so-called A_z value is the area under the ROC curve, and the A_z value is a concise description of the diagnostic performance of the systems being tested. The A_z is a good measure of what may be called *detectability*, recognizing that for worst performance $A_z = 0.5$ and for best performance, $A_z = 1.0$.

SUGGESTED READING

- Barrett HH, Swindell W. *Radiological imaging: the theory of image formation, detection, and processing*, vols 1 and 2. New York: Academic Press, 1981.
- Bracewell RN. *The Fournier transform and its applications*. New York: McGraw-Hill, 1978.
- Dainty JC, Shaw R. *Image science*. New York: Academic Press, 1974.
- Hasegawa BH. *The physics of medical x-ray imaging*, 2nd ed. Madison, WI: Medical Physics, 1991.
- Rose A. *Vision: human and electronic*. New York: Plenum Press, 1973.

DIGITAL RADIOGRAPHY

Digital radiographic image receptors are gradually replacing screen-film cassettes as radiology departments convert to an all-digital environment. Computed tomography (CT) and magnetic resonance imaging (MRI) are intrinsically digital, and ultrasound and nuclear medicine imaging made the change to digital imaging from their analog ancestries in the 1970s. Thus, radiography is really the last modality to make the transition to digital acquisition. The reasons for this are clear: Screen film is a tried and true detector system that produces excellent image quality under most circumstances, and therefore the motivation for change is low. In addition, the large field of view and high spatial resolution of radiography require digital radiographic images to contain large amounts of digital data. For example, a typical high-resolution digital chest radiograph ranges from 4 megabytes (MB) ($2\text{ k} \times 2\text{ k} \times 8\text{ bit}$) to 32 MB ($4\text{ k} \times 4\text{ k} \times 12\text{ bit}$). By comparison, a single CT image is 0.5 MB, and a single photon emission computed tomography (SPECT) image is just 16 kilobytes (kB). The disadvantages of the large size of digital radiographic images are threefold: (a) they require lots of space on digital storage media, (b) they require high network bandwidth in a picture archiving and communication system (PACS), and (c) they require costly high luminance and high resolution monitors (typically $2\text{ k} \times 2.5\text{ k}$, if entire images are to be displayed at full or almost full spatial resolution) for display.

11.1 COMPUTED RADIOGRAPHY

Computed radiography (CR) is a marketing term for photostimulable phosphor detector (PSP) systems. Phosphors used in screen-film radiography, such as $\text{Gd}_2\text{O}_2\text{S}$ emit light promptly (virtually instantaneously) when struck by an x-ray beam. When x-rays are absorbed by photostimulable phosphors, some light is also promptly emitted, but much of the absorbed x-ray energy is trapped in the PSP screen and can be read out later. For this reason, PSP screens are also called *storage phosphors* or *imaging plates*. CR was introduced in the 1970s, saw increasing use in the late 1980s, and was in wide use at the turn of the century as many departments installed PACS, often in concert with the development of the electronic medical record.

CR imaging plates are made of BaFBr and BaFI. Because of this mixture, the material is often just called barium fluorohalide. A CR plate is a flexible screen that is enclosed in a cassette similar to a screen-film cassette. One imaging plate is used

for each exposure. The imaging plate is exposed in a procedure identical to screen-film radiography, and the CR cassette is then brought to a CR reader unit. The cassette is placed in the readout unit, and several processing steps then take place:

1. The cassette is moved into the reader unit and the imaging plate is mechanically removed from the cassette.
2. The imaging plate is translated across a moving stage and scanned by a laser beam.
3. The laser light stimulates the emission of trapped energy in the imaging plate, and visible light is released from the plate.
4. The light released from the plate is collected by a fiber optic light guide and strikes a photomultiplier tube (PMT), where it produces an electronic signal.
5. The electronic signal is digitized and stored.
6. The plate is then exposed to bright white light to erase any residual trapped energy.
7. The imaging plate is then returned to the cassette and is ready for reuse.

The digital image that is generated by the CR reader is stored temporarily on a local hard disk. Many CR systems are joined ("docked") directly to laser printers that make film hard copies of the digital images. CR systems often serve as entry points into a PACS, and in such cases the digital radiographic image is sent to the PACS system for interpretation by the radiologist and long-term archiving.

The imaging plate is a completely analog device, but it is read out by analog and digital electronic techniques, as shown in Fig. 11-1. The imaging plate is translated along the readout stage in the vertical direction (the y direction), and a scanning laser beam interrogates the plate horizontally (the x direction). The laser is scanned with the use of a rotating multifaceted mirror. As the red laser light (approximately 700 nm) strikes the imaging phosphor at a location (x,y), the trapped energy from the x-ray exposure at that location is released from the imaging plate. A fraction of the emitted light travels through the fiberoptic light guide

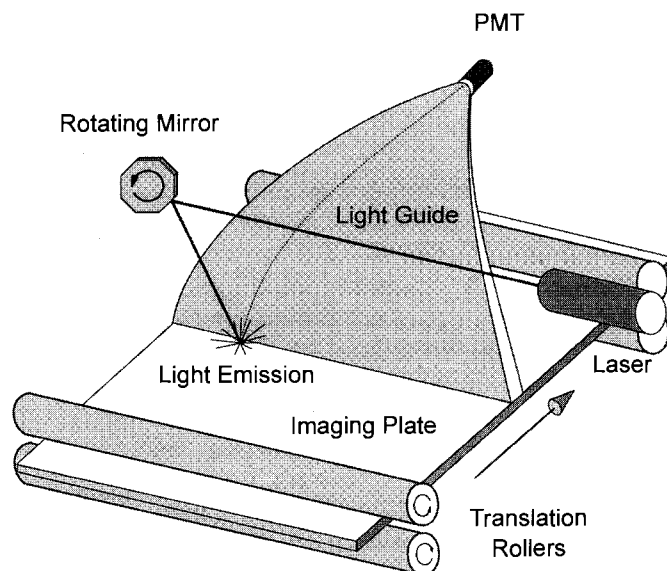


FIGURE 11-1. The hardware of a computed radiography (CR) reader is illustrated. The imaging plate is mechanically translated through the readout system using rollers. A laser beam strikes a rotating multifaceted mirror, which causes the beam to repeatedly scan across the imaging plate. These two motions result in a raster scan pattern of the imaging plate. Light is emitted from the plate by laser stimulation. The emitted light travels through a fiberoptic light guide to a photomultiplier tube (PMT), where the light is converted to an electronic signal. The signal from the PMT is subsequently digitized.

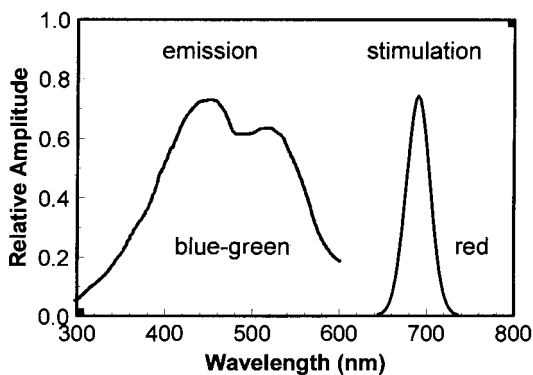


FIGURE 11-2. The optical spectra used in a computed radiography (CR) system are illustrated. The red laser light stimulates the release of trapped x-ray energy (electrons trapped in excited states) in the imaging plate. When the trapped electron energy is released, a broad spectrum of blue-green light is emitted. An optical filter, placed in front of the photomultiplier tube (PMT), prevents the detection of the red laser light.

and reaches a PMT (discussed in Chapter 20). The electronic signal that is produced by the PMT is digitized and stored in memory. Therefore, for every spatial location (x,y) , a corresponding gray scale value is determined, and this is how the digital image $I(x,y)$ is produced in a CR reader.

The light that is released from the imaging plate is of a different color than the stimulating laser light (Fig. 11-2). To eliminate detection by the PMT of the scattered laser light, an optical filter that transmits the blue-green light that is emitted from the imaging plate, but attenuates the red laser light, is mounted in front of the PMT.

Figure 11-3 describes how stimutable phosphors work. A small mass of screen material is shown being exposed to x-rays. Typical imaging plates are composed of

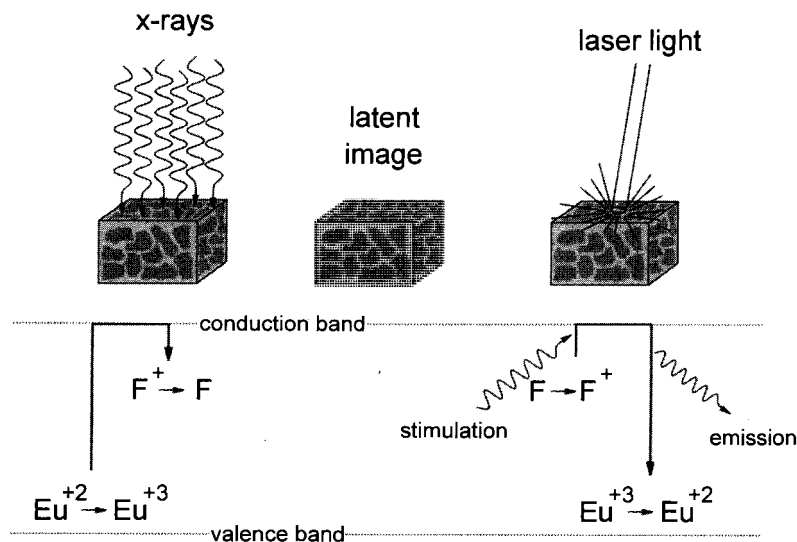


FIGURE 11-3. This figure illustrates the sequence of events during the x-ray exposure and readout of a photostimulable phosphor. During x-ray exposure to the imaging plate, electrons are excited from the valence band to the conduction band, and some of these are trapped in the F-centers. After x-ray exposure, the latent image exists as a spatially dependent distribution of electrons trapped in high-energy states. During readout, laser light is used to stimulate the trapped electrons back up to the conduction band, where they are free to transition to the valence band, and in doing so blue-green light is emitted. The light intensity emitted by the imaging plate is proportional to the absorbed x-ray energy.

about 85% BaFBr and 15% BaFI, activated with a small quantity of europium. As discussed in Chapter 6, the nomenclature BaFBr:Eu indicates the BaFBr phosphor activated by europium. This activation procedure, also called doping, creates defects in the BaFBr crystals that allow electrons to be trapped more efficiently.

When the x-ray energy is absorbed by the BaFBr phosphor, the absorbed energy excites electrons associated with the europium atoms, causing divalent europium atoms (Eu^{+2}) to be oxidized and changed to the trivalent state (Eu^{+3}). The excited electrons become mobile, and some fraction of them interact with a so-called F-center. The F-center traps these electrons in a higher-energy, metastable state, where they can remain for days to weeks, with some fading over time. The latent image that exists on the imaging plate after x-ray exposure, but before readout, exists as billions of electrons trapped in F-centers. The number of trapped electrons per unit area of the imaging plate is proportional to the intensity of x-rays incident at each location during the exposure.

When the red laser light scans the exposed imaging plate, the red light is absorbed at the F-center, where the energy of the red light is transferred to the electron. The photon energy of the red laser light is less than that of the blue-green emission ($\Delta E_{\text{red}} < \Delta E_{\text{blue-green}}$). However, the electron gains enough energy to reach the conduction band, enabling it to become mobile again. Many of these electrons then become de-excited by releasing blue-green light as they become reabsorbed by the trivalent europium atoms, converting them back to the divalent state (see Fig. 11-3). This is how the red laser light stimulates emission of the blue and green light photons from the imaging plate.

The first readout of the imaging plate does not release all of the trapped electrons that form the latent image; indeed, a plate can be read a second time and a third time with only slight degradation. To erase the latent image so that the imaging plate can be reused for another exposure without ghosting, the plate is exposed

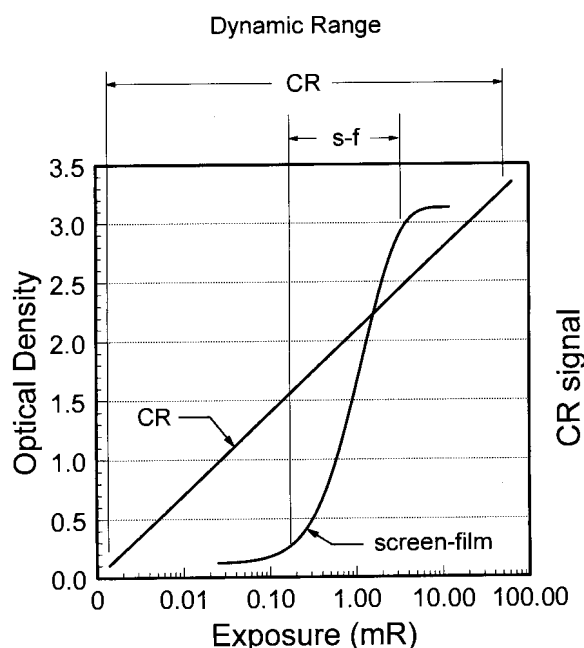


FIGURE 11-4. One of the principal advantages of computed radiography (CR) over screen-film detector systems is the greater dynamic range that CR provides. The exposure range (film latitude) in which screen-film cassettes produce usable optical densities is illustrated. The usable exposure range for the CR system is much greater.

to a very bright light source, which flushes almost all of the metastable electrons to their ground state, emptying most of the F-centers.

CR systems operate with a workflow pattern very similar to screen-film radiography. CR cassettes are exposed just like screen-film cassettes, and they are then brought to a reader unit, just as film-screen cassettes are brought to the film processor. The similarity in handling of CR and screen-film cassettes contributed to the initial success of CR. One of the advantages that CR over screen-film radiography is its much larger dynamic range (Fig. 11-4); in other words, the exposure latitude with CR is much wider than with screen-film systems. Consequently, retakes due to overexposure or underexposure are rare with CR. Because of the difficulty of obtaining precise exposures in the portable radiographic setting, where phototiming usually is not available, CR is often used for portable examinations. Although CR is capable of producing images with proper gray scale at high and low exposure levels, quality assurance efforts should still be in place to ensure proper exposure levels. CR images exposed to low exposure levels, while maintaining proper gray scale in the image, have higher levels of x-ray quantum noise ("radiographic mottle"). CR images produced at high exposures have low quantum noise but result in higher x-ray dose to the patient.

11.2 CHARGED-COUPLED DEVICES

Charged-coupled devices (CCDs) form images from visible light. CCD detectors are used in most modern videocameras and in digital cameras. The principal feature of CCD detectors is that the CCD chip itself is an integrated circuit made of crystalline silicon, similar to the central processing unit of a computer. A CCD chip has discrete pixel electronics etched into its surface; for example, a 2.5×2.5 cm CCD chip may have 1024×1024 or 2048×2048 pixels on its surface. Larger chips spanning 8×8 cm have been produced. The silicon surface of a CCD chip is photo-sensitive—as visible light falls on each pixel, electrons are liberated and build up in the pixel. More electrons are produced in pixels that receive greater light intensity. The electrons are kept in each pixel because there are electronic barriers (voltage) on each side of the pixel during exposure.

Once the CCD chip has been exposed, the electronic charge that resides in each pixel is read out. The readout process is akin to a bucket brigade (Fig. 11-5). Along one column of the CCD chip, the electronic charge is shifted pixel by pixel by appropriate control of voltage levels at the boundaries of each pixel. The charge from each pixel in the entire column is shifted simultaneously, in parallel. For linear CCD detectors, the charge packet that is at the very bottom of the linear array spills onto a transistor, where it produces an electronic signal that is digitized. The entire line of pixels is thus read out one by one, in a shift-and-read, shift-and-read process. For a two-dimensional CCD detector, the charges on each column are shifted onto the bottom row of pixels, that entire row is read out horizontally, and then the next charges from all columns are shifted down one pixel, and so on.

CCD cameras produce high-quality images from visible light exposure and are commonly used in medical imaging for fluoroscopy and digital cineradiography. In these devices, the amplified light generated by the image intensifier is focused with the use of lenses or fiberoptics onto the CCD chip (Fig. 11-6). For small field-of-view applications such as dental radiography (e.g., 25×50 mm detector), an inten-

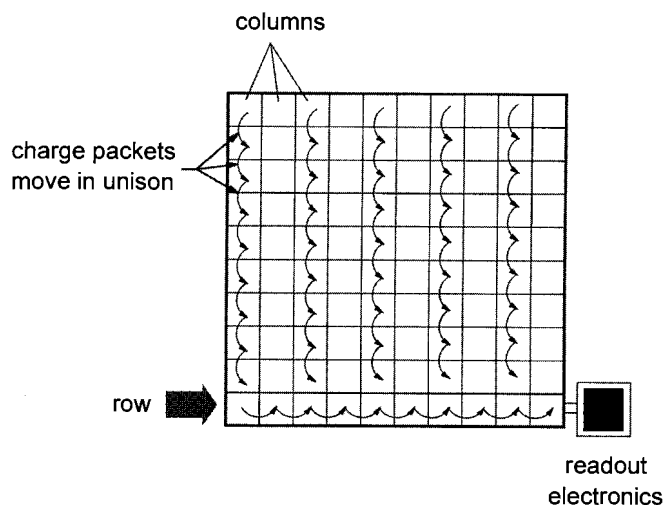


FIGURE 11-5. Charged-coupled devices (CCDs) are pixelated light detectors in which electronic charge builds up in each detector element as a response to light exposure. After exposure, by toggling voltages between adjacent detector elements, the charge packets are made to move in unison (simultaneously) down each column of the array. After the charge packets have been shifted down one detector element in each column, the bottom row of pixels is refreshed. This bottom row is read out horizontally, in a bucket brigade fashion, so that the charge packet from each detector element reaches the readout electronics of the CCD. After the entire bottom row is read out, the charge packets from all columns are again shifted down one pixel, refreshing the bottom row for subsequent readout. This process is repeated until the signal from each detector element has been read out and digitized, producing the digital image.

sifying screen is placed in front of the CCD chip and the system is exposed. The light emitted from the screen is collected by the CCD chip, and because of the excellent coupling between the screen and the CCD chip (which are in contact with each other) only a relatively small amount of the light generated in the screen is wasted (i.e., does not reach the CCD). For applications in which the field of view is only slightly larger than the area of the CCD chip, such as in digital biopsy systems for mammography, a fiberoptic taper is placed between the intensifying screen and the CCD (see Fig. 11-6). The fiberoptic taper serves as an efficient lens that focuses the light produced in the screen onto the surface of the CCD chip. Typical fiberoptic tapers used in digital biopsy have front surfaces that measure 50×50 mm (abutted to the screen) and output surfaces that are 25×25 mm (abutted to the CCD). In these systems, the amount of light lost in the fiberoptic taper is not trivial, but a substantial fraction of the light reaches the CCD chip.

When there is a need to image a large field of view, such as in chest radiography, it is impossible to focus the light emitted from the large screen (approximately 35×43 cm) onto the surface of the CCD chip without losing a very large fraction of the light photons. The amount of light lost is proportional to the demagnification factor required to couple the input area to the output area. As an example, the area ratio (demagnification factor) for 35×43 -cm chest radiography using a 30×30 -mm CCD chip is 167:1. Even with perfect lenses in such a system, the light loss

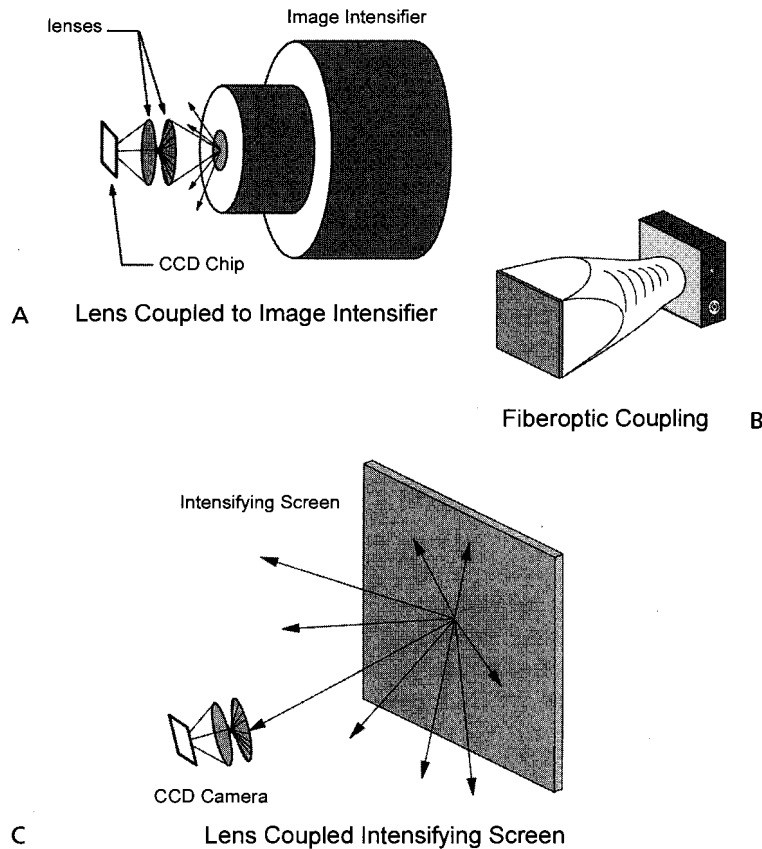


FIGURE 11-6. A: The top frame shows a charged-coupled device (CCD) chip coupled through tandem lenses to the output of an image intensifier. Because of the huge light amplification of the image intensifier, many thousands of light photons are produced by the absorption of one x-ray photon. A small fraction of these light photons pass through the lenses and are focused onto the CCD chip. Although many light photons are lost in the lens and aperture, the huge gain of the image intensifier ensures that a secondary quantum sink is avoided. **B:** Digital mammography biopsy systems make use of fiberoptic coupling between a small field-of-view intensifying screen and a CCD chip. Because the demagnification factor is low and the fiberoptic lens is efficient, a secondary quantum sink is avoided. **C:** For large field-of-view applications, the use of lenses coupled to an intensifying screen results in a substantial loss of the light photons emitted by the x-ray interaction in the screen. This type of system suffers from a secondary quantum sink.

is approximately 99.7% (see Fig. 11-6). As discussed in Chapter 10, if an insufficient number of quanta (light photons) are used to conduct an image through an imaging chain after the initial x-ray detection stage, a secondary quantum sink will occur. Although an image can be produced if a sufficient number of x-ray photons are used, any x-ray system with a secondary quantum sink will have image quality that is not commensurate with the x-ray dose used to make the image.

11.3 FLAT PANEL DETECTORS

Flat panel detector systems make use of technology similar to that used in a laptop computer display, and much of this has to do with wiring the huge number of individual display elements. Instead of producing individual electrical connections to each one of the elements in a flat panel display, a series of horizontal and vertical electrical lines are used which, when combined with appropriate readout logic, can address each individual display element. With this approach only 2,000 connections between the imaging plate and the readout electronics are required for a $1,000 \times 1,000$ display, instead of 1,000,000 individual connections. For a flat panel display, the wiring is used to send signals from the computer graphics card to each display element, whereas in a detector the wiring is used to measure the signal generated in each detector element.

Indirect Detection Flat Panel Systems

Indirect flat panel detectors are sensitive to visible light, and an x-ray intensifying screen (typically $\text{Gd}_2\text{O}_2\text{S}$ or CsI) is used to convert incident x-rays to light, which is then detected by the flat panel detector. The term “indirect” comes from the fact that x-rays are absorbed in the screen, and the absorbed x-ray energy is then relayed to the photodetector by visible light photons. This indirect detection strategy is analogous to a screen-film system, except that the electronic sensor replaces the light-sensitive film emulsion. Dual-emulsion film is thin and x-rays penetrate it easily; therefore, in a screen-film cassette, the film is sandwiched between screens to reduce the average light propagation path length and improve spatial resolution. Flat panels are thicker than film and do not transmit x-rays well; consequently, a sandwiched design is not possible. Instead, the intensifying screen is layered on the front surface of the flat panel array. This means that the light emanating from the *back* of the intensifying screen strikes the flat panel, and much of the light that is released in the screen has to propagate relatively large distances through the screen, which results in more blurring. To improve this situation, most flat panel detector systems for general radiography use CsI screens instead of $\text{Gd}_2\text{O}_2\text{S}$. CsI is grown in columnar crystals (as described in Chapter 9), and the columns act as light pipes to reduce the lateral spread of light.

A typical configuration for a flat panel detector system is shown in Fig. 11-7. The flat panel comprises a large number of individual detector elements, each one capable of storing charge in response to x-ray exposure. Each detector element has a light-sensitive region, and a small corner of it contains the electronics. Just before exposure, the capacitor, which stores the accumulated x-ray signal on each detector element, is shunted to ground, dissipating lingering charge and resetting the device. The light-sensitive region is a photoconductor, and electrons are released in the photoconductor region on exposure to visible light. During exposure, charge is built up in each detector element and is held there by the capacitor. After exposure, the charge in each detector element is read out using electronics as illustrated in Fig. 11-8.

A transistor is a simple electronic switch that has three electrical connections—the gate, the source, and the drain. Each detector element in a flat panel detector has a transistor associated with it; the source is the capacitor that stores the charge accumulated during exposure, the drain is connected to the readout line (vertical

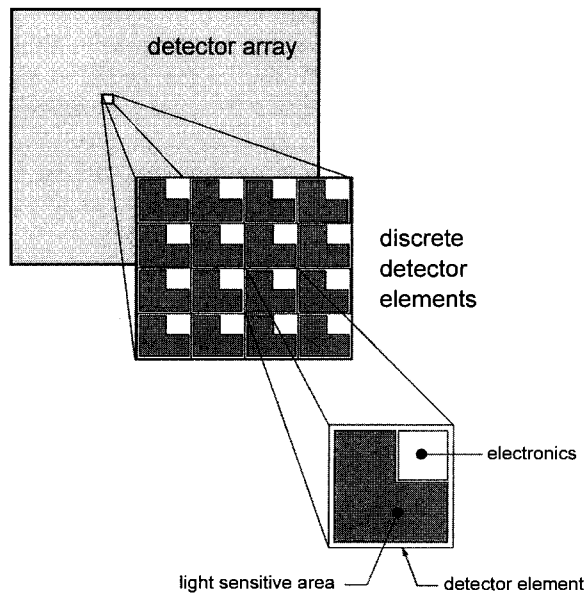


FIGURE 11-7. A flat panel detector array comprises a large number of discrete detector elements. Each detector element contains both a light-sensitive area and a region that holds electronic components.

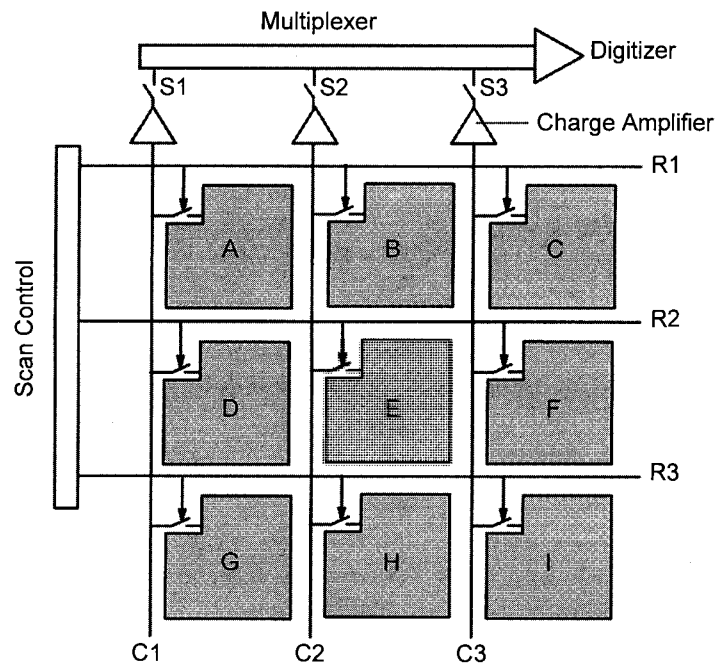


FIGURE 11-8. This figure illustrates the readout process used for flat panel detector arrays. Nine detector elements are shown (A through I). Three gate lines (rows R1, R2, and R3) and three readout lines (columns C1, C2, and C3) are illustrated.

wires in Fig. 11-8), and the gate is connected to the horizontal wires shown in Fig. 11-8. Negative voltage applied to the gate causes the switch to be turned *off* (no conduction from source to drain), whereas when a positive voltage is applied to the gate the switch is turned *on* (source is connected to drain). Because each detector element has a transistor and the device is manufactured using thin-film deposition technology, these flat panel systems are called thin-film transistor (TFT) image receptors.

The readout procedure occurs as follows. During exposure, negative voltage is applied to all gate lines, causing all of the transistor switches on the flat panel imager to be turned off. Therefore, charge accumulated during exposure remains at the capacitor in each detector element. During readout, positive voltage is sequentially applied to each gate line (e.g., R1, R2, R3, as shown in Fig. 11-8), one gate line at a time. Thus, the switches for all detector elements along a row are turned on. The multiplexer (top of Fig. 11-8) is a device with a series of switches in which one switch is opened at a time. The multiplexer sequentially connects each vertical wire (e.g., C1, C2, C3), via switches (S1, S2, S3), to the digitizer, allowing each detector element along each row to be read out. For example, referring to Fig. 11-8, when wire R2 is set to a positive gate voltage (all other horizontal wires being negative), the switches on detector elements D, E, and F are opened. Therefore, current can in principle flow between each of these detector elements (source) and the digitizer (drain). In the multiplexer, if the switches S1 and S3 are turned off and S2 is on, then the electrons accumulated on detector element E are free to flow (under the influence of an applied voltage) from the storage capacitor, through the charge amplifier, through the multiplexer (S2 is open) to the digitizer. Thus, the array of detector elements is read out in a raster fashion, with the gate line selecting the row and the multiplexer selecting the column. By this sequential readout approach, the charge from each detector element is read out from the flat panel, digitized, and stored, forming a digital image. Notice that in this procedure the signal from each detector element does not pass through any other detector elements, as it does in a CCD camera. This means that the charge transfer efficiency in a flat panel imager needs only to be good (approximately 98%), whereas for a CCD the charge transfer efficiency needs to be excellent (greater than 99.99%). Therefore, flat panel systems are less susceptible to imperfections that occur during fabrication. This translates into improved manufacturing cost efficiency.

The size of the detector element on a flat panel largely determines the spatial resolution of the detector system. For example, for a flat panel with $125 \times 125\text{-}\mu\text{m}$ pixels, the maximum spatial frequency that can be conveyed in the image (the *Nyquist frequency*, F_N) is $(2 \times 0.125\text{ mm})^{-1}$, or 4 cycles/mm. If the detector elements are $100\text{ }\mu\text{m}$, then $F_N = 5\text{ cycles/mm}$. Because it is desirable to have high spatial resolution, small detector elements are needed. However, the electronics (e.g., the switch, capacitor, etc.) of each detector element takes up a certain (fixed) amount of the area, so, for flat panels with smaller detector elements, a larger fraction of the detector element's area is not sensitive to light. Therefore, the light collection efficiency decreases as the detector elements get smaller. The ratio of the light-sensitive area to the entire area of each detector element is called the *fill factor* (Fig. 11-9). It is desirable to have a high fill factor, because light photons that are not detected do not contribute to the image. If a sufficient number of the light photons generated in the intensifying screen are lost owing to a low fill factor, then contrast resolution (which is related to the signal-to-noise ratio) will be degraded. Therefore, the choice

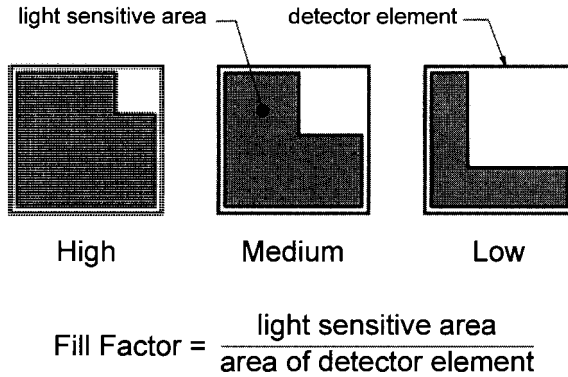


FIGURE 11-9. For indirect detection flat panel arrays, the light collection efficiency of each detector element depends on the fractional area that is sensitive to light. This fraction is called the *fill factor*, and a high fill factor is desirable.

of the detector element dimensions requires a tradeoff between spatial resolution and contrast resolution.

Direct Detection Flat Panel Systems

Direct flat panel detectors are made from a layer of photoconductor material on top of a TFT array. These photoconductor materials have many of the same properties as silicon, except they have higher atomic numbers. Selenium is commonly used as the photoconductor. The TFT arrays of these direct detection systems make use of the same line-and-gate readout logic, as described for indirect detection systems. With direct detectors, the electrons released in the detector layer from x-ray interactions are used to form the image directly. Light photons from a scintillator are not used. A negative voltage is applied to a thin metallic layer (electrode) on the front surface of the detector, and therefore the detector elements are held positive in relation to the top electrode. During x-ray exposure, x-ray interactions in the selenium layer liberate electrons that migrate through the selenium matrix under the influence of the applied electric field and are collected on the detector elements. After exposure, the detector elements are read out as described earlier for indirect systems.

With indirect systems, the light that is released in the intensifying screen diffuses as it propagates to the TFT array, and this causes a certain amount of blurring (Fig. 11-10). This blurring is identical to that experienced in screen-film detectors. For indirect flat panel detectors, the columnar structure of CsI is often exploited to help direct the light photons toward the detector and reduce lateral spread. This approach works well, but a significant amount of lateral blurring does still occur. For direct detection systems, in contrast, electrons are the secondary quanta that carry the signal. Electrons can be made to travel with a high degree of directionality by the application of an electric field (see Fig. 11-10). Therefore, virtually no blurring occurs in direct detection flat panel systems. Furthermore, electric field lines can be locally altered at each detector element by appropriate electrode design, so that the sensitive area of the detector element collects electrons that would otherwise reach insensitive areas of each detector element, increasing the effective fill factor. Because of the ability to direct the path of electrons in direct detection flat panel systems, the spatial resolution is typically limited only by the dimensions of the detector element.

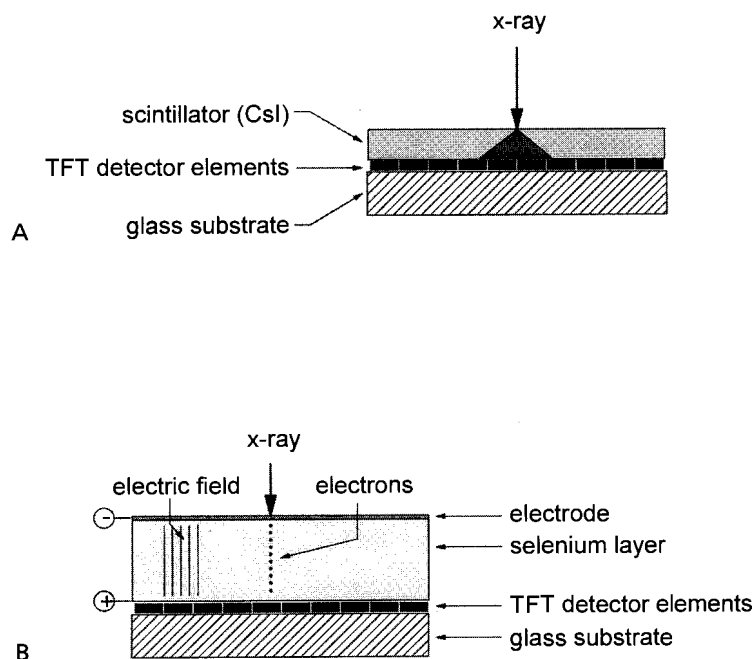


FIGURE 11-10. A: A cross-section of an indirect detection flat panel system is shown. X-rays interact with a scintillator, where visible light is emitted and diffuses toward the photosensitive detector elements of the panel. **B:** Direct flat panel detectors often use a layer of amorphous selenium photoconductor coupled to a thin-film transistor (TFT) array. In a direct detection flat panel device, x-ray interaction in the selenium layer releases electrons (or electron holes), which are used to form the signal directly. An electric field applied across the selenium layer minimizes the lateral spread of the electrons, preserving spatial resolution.

Selenium ($Z = 34$) has a higher atomic number than silicon ($Z = 14$), but it is still quite low compared with conventional x-ray intensifying screen phosphors that contain elements such as gadolinium ($Z = 64$) or cesium ($Z = 55$), and therefore the attenuation coefficient is relatively low at diagnostic x-ray energies (40 to 130 keV). Intensifying screens need to be kept thin enough to reduce light spreading and thereby preserve spatial resolution. However, because electrons do not spread laterally in selenium direct detectors due to the electric field, selenium detectors are made much thicker than indirect detection systems, to improve x-ray detection efficiency. Making the selenium detectors thicker compensates for the relatively low x-ray absorption of selenium. In addition to selenium, other materials such as mercuric iodide (HgI_2), cadmium telluride (CdTe), and lead iodide (PbI_2) are being studied for use in direct detection flat panel systems.

11.4 DIGITAL MAMMOGRAPHY

Mammography presents some of the greatest challenges for digital detectors because of the high spatial resolution that are required to detect and characterize microcalcifications. Small field-of-view digital mammography has been used for several years

for digital stereotactic biopsy. Breast biopsy systems using digital detectors are typically 5×5 cm and make use of either fiberoptic tapers or mirrors coupled to CCD cameras (Fig. 11-11). X-ray energy absorbed in the intensifying screen releases light, which is conducted through light pipes that focus the image onto the surface of a CCD chip. A secondary quantum sink is avoided because the fiberoptic lens is quite efficient and the demagnification factor is small.

Full-field digital mammography systems based on CCD cameras make use of a mosaic of CCD systems. The use of several CCD chips is necessary to increase the area of the light sensor and thereby keep the demagnification factor low (similar to that of digital biopsy) in order to avoid a secondary quantum sink. A mosaic CCD system based on fiberoptic tapers is shown in Fig. 11-11, and lens-coupled systems (still with low demagnification factors) exist as well. With mosaic CCD systems, the

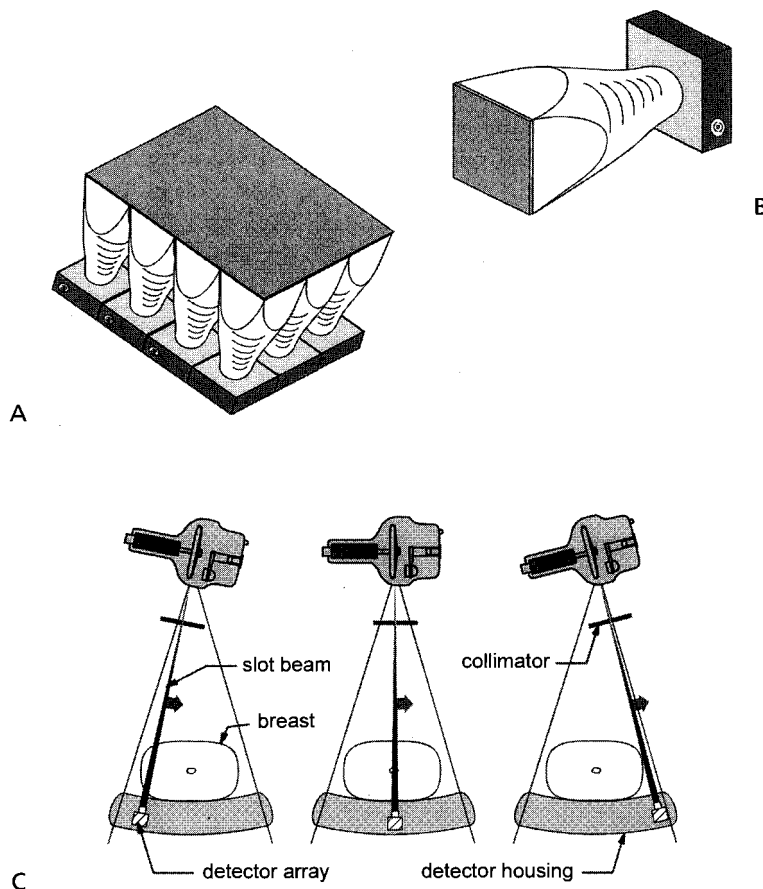


FIGURE 11-11. A: A digital biopsy system is illustrated. The system uses a fiberoptic taper to couple a small intensifying screen to the chip of a Charged-coupled device (CCD). **B:** By combining several digital breast biopsy modules into a mosaic, a full-field digital mammography detector can be assembled. **C:** Image acquisition using a slot-scan system for digital mammography is illustrated. A narrow slot beam of x-rays scans across the breast, striking a CCD detector system. The detector system moves inside a plastic detector housing.

individual images from each camera are *stitched* together with the use of software to produce a seamless, high-resolution digital mammographic image.

Full-field mammography systems (18×24 cm) are available that utilize TFT/flat panel detector technology. Currently, indirect detection (CsI scintillator) flat panel array systems are available with $100 \times 100\text{-}\mu\text{m}$ detector elements ($F_N = 5$ cycles/mm); however, flat panels (indirect and direct) with smaller detector elements are being developed. Flat panel mammography devices are high-resolution versions (thinner screens and smaller detector elements) of standard flat panel systems, which were described previously.

Another technology that is used for full-field digital mammography is the slot-scan system (see Fig. 11-11). The slot-scan system uses a long, narrow array of CCD chips. Because the largest CCD arrays are only 6- to 8-cm square, full-field mammography using a single CCD is not possible. However, the slot-scan system requires a detector array with dimensions of about $4\text{ mm} \times 18\text{ cm}$, and several long, narrow CCD chips can meet this need. The breast remains stationary and under compression while the x-ray beam is scanned across it. At any time during the acquisition sequence, only a narrow slot of tissue is being exposed to the highly collimated x-ray beam. Consequently, little scattered radiation is detected due to the narrow beam geometry. The geometric efficiency of the slot-scan system is quite low compared with that of full-field systems, because only the x-rays that pass through the narrow collimation into the slot are used to make the image. To acquire images in clinically realistic time frames (1 to 2 seconds), the x-ray tube in this system uses a tungsten anode and is operated at higher peak kilovoltage (kVp) than in conventional mammography (about 45 kV). Tungsten is a more efficient anode material than molybdenum, and x-ray emission is more efficient at higher kVp. These two factors help overcome the heat-loading issues that can be a problem with slot-scan geometries.

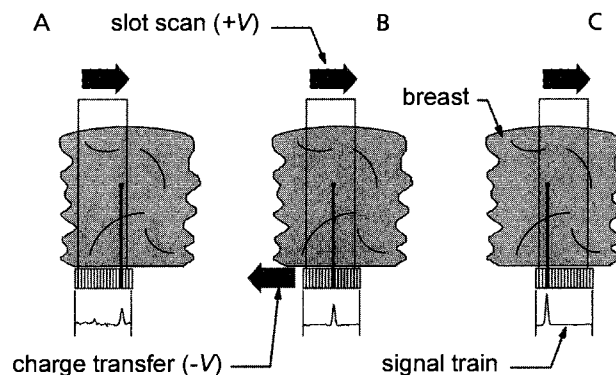


FIGURE 11-12. The readout logic of a time delay and integration (TDI) system, which is used for slot-scan digital mammography, is illustrated. The breast remains stationary under compression during image formation. **A** through **C**: The x-ray beam is collimated by a slot (as shown), and the slot is scanned across the breast. While the x-ray beam is scanned from left to right at velocity $+V$, the CCD is read out in the opposite direction at velocity $-V$. The x-ray shadow of a small structure in the breast (**dark circle**) projects onto the same position of the signal train on the CCD camera as the signal train moves across the chip. Because of the synchrony between the slot scan and the CCD readout, the image of fine structures in the breast is preserved with high resolution.

Because the slot width (4 mm) is much wider than one pixel (0.050 mm), scanning the system with the detector moving, as shown in Fig. 11-11, would result in image blurring if the detector motion were not compensated for. In Fig. 11-12 a slot-scan system is shown moving from left to right with a velocity $+V$. To exactly compensate for this motion, the CCD chip is read out in bucket brigade fashion in the direction opposite to the scan motion, at the same velocity $(-V)$. With this *time delay and integration* (TDI) readout approach, the x-ray shadow of an object in the breast is cast onto one of the CCD detector elements (see Fig. 11-12A). As the scanning x-ray beam and detector move to the right (see Fig. 11-12B), the charge packets in the CCD array are moving to the left. Therefore, from the time a tiny object enters the slot to the time it exits the slot, its x-ray shadow continually falls at the same position along the electronic signal train as it moves across the chip from right to left. This compensatory motion essentially freezes the motion of the detected x-ray information, allowing the spatial resolution across the slot width to be dictated by the CCD pixel size instead of by the slot width itself. This trick increases the sensitivity of the scan-slot detector system (and thus reduces tube loading) by the ratio of the slot width (4 mm) to the pixel width (0.050 mm), a substantial factor of 80.

11.5 DIGITAL VERSUS ANALOG PROCESSES

Although image receptors are referred to as “digital,” the initial stage of these devices produces an analog signal. In CR, a photomultiplier tube detects the visible light emitted from the photostimulable phosphor plate and produces a current that is subsequently digitized. Although CCD and flat panel detector systems are divided into discrete detector elements (and therefore are referred to as “pixelated” detectors), the signal produced in each pixel is an analog packet of electronic charge which is digitized by an analog-to-digital converter (ADC) during image readout (see Chapter 4). Therefore, the initial stages of all digital detectors involve analog signals.

11.6 IMPLEMENTATION

The choice of a digital detector system requires a careful assessment of the needs of the facility. CR is often the first digital radiographic system that is installed in a hospital, because it can be used for providing portable examinations, where the cost of retakes in both time and money is high. A significant benefit of CR over TFT-based digital radiographic technology is that the reader is stationary, but because the CR plates themselves are relatively inexpensive, several dozen plates may be used in different radiographic rooms simultaneously. The number of rooms that one CR reader can serve depends on the types of rooms, workload, and proximity. Typically, one CR reader can handle the workload of three radiographic rooms. The cost of an individual flat panel detector is high, and it can only be in one place at one time. The obvious utility for flat panel detectors is for radiographic suites that have high patient throughput, such as a dedicated chest room. Because no cassette handling is required and the images are rapidly available for checking positioning, the throughput in a single room using flat panel detectors can be much higher than for screen-film or CR systems.

11.7 PATIENT DOSE CONSIDERATIONS

When film-screen image receptors are used, an inadvertent overexposure of the patient will result in a dark film, which provides immediate feedback to the technologist regarding the technique factors (and relative dose) used. However, when digital image receptors are used, overexposure of the patient can produce excellent images because the electronic systems compensate for (and essentially mask) large fluctuations in exposure. Consequently, high-exposure conditions may go unnoticed unless appropriate measures for periodic monitoring are taken. For this reason, a quality control program should be in place to ensure proper exposure levels. Most digital detector systems provide an estimate of the incident exposure that can assist in the evaluation of exposure trends. The exposures necessary to produce good images are directly related to the detective quantum efficiency (DQE) of the detector: Detectors with high DQEs make more efficient use of the x-rays and therefore require less exposure for adequate signal-to-noise ratios. X-ray exposure levels should be tailored to the needs of the specific clinical examination, in consideration of the digital detector and its DQE. For example, it is generally accepted that CR imaging systems require about twice the exposure of a corresponding 400-speed screen-film detector for comparable image quality. The CR system is therefore equivalent to a 200-speed screen-film system for most general radiographic examinations. Screen-film cassettes for extremity radiography are slower (i.e., require more radiation) than those used for general radiography, and higher exposure levels are required for extremity imaging with CR as well (equivalent to 75- to 100-speed screen-film systems).

Flat panel detectors can reduce radiation dose by about twofold to threefold for adult imaging, compared with CR for the same image quality, owing to the better quantum absorption and conversion efficiency associated with that technology.

11.8 HARD COPY VERSUS SOFT COPY DISPLAY

Hard copy display refers to displaying images on film, and *soft copy* display refers to using video monitors. Digital radiographic images produced by flat panel TFT arrays or by CR systems are significantly larger in matrix size than any other images produced in the radiology department. For example, typical SPECT and positron emission tomography (PET) images are 128×128 pixels, MRI images are typically 256×256 pixels, and CT images are 512×512 pixels. Digital subtraction angiography images are often $1,024 \times 1,024$ pixels. Digital chest radiographs using 100- μm pixels result in an image matrix of $3,500 \times 4,300$ pixels. A typical personal computer display screen is slightly larger than 1×1 k, and therefore could simultaneously display 64 full-resolution SPECT images, 16 full-resolution MRI images, or 4 full-resolution CT images. However, the 3.5×4.3 k-pixel digital chest image is much too large for a 1×1 k monitor; only one-fifteenth of the image data could be displayed at any one time. Therefore, when digital radiographs are included in the image mix, the display resolution of the monitors used for soft copy reading must be increased if entire images are to be displayed at full or nearly full spatial resolution. This increases the cost of soft copy display workstations. A more detailed discussion of soft copy display is given in Chapter 17.

11.9 DIGITAL IMAGE PROCESSING

One of the advantages of having an image in digital format is that its appearance can be modified and sometimes improved using a computer to manipulate the image. Image processing is a vast subject on which many books have been written. Here, a brief introduction of image processing is given.

Digital Image Correction

Most digital radiographic detector systems have an appreciable number of defects and imperfections. There is even a grading system for CCD chips that relates to the number of defects in the chip. Fortunately, under most situations the blemishes in the image can be corrected using software. For both CCD and flat panel systems, invariably there are a certain number of detector elements that do not work—they generate no response to light input and are called “dead pixels.” After manufacture, each detector is tested for dead pixels, and a dead pixel map is made for each individual detector array or CCD chip. Software uses the dead pixel map to correct the images produced by that detector: The gray scale values in the pixels surrounding the dead pixel are averaged, and this value replaces the gray scale of the dead pixel. When the fraction of dead pixels is small, the correction procedure makes it almost impossible to see dead pixel defects in the image. Because of the bucket brigade type of readout that a CCD uses, some dead pixels result in the loss of a column of data (i.e., when charge cannot be transferred through the dead pixel). Column defects can also be corrected with the use of interpolation procedures; however, different interpolation techniques are used for column defects than for dead pixels.

In the absence of x-rays, each detector element has a certain amount of electronic noise associated with it (“dark noise”). Finite gray scale values result from image acquisition in the absence of x-rays, depending on the dark noise in each individual detector element. These noise values are different from exposure to exposure, but a large number of dark images can be acquired and averaged together, generating an *average* dark image $D(x,y)$. This dark image is subtracted from the images produced by the detector array.

Even for properly functioning detectors, there are subtle differences in the sensitivity of each detector element. For large detector arrays, several amplifiers are used, and each one has a slightly different gain. In some digital detector array systems, the amplifiers are associated with different regions (i.e., adjacent groups of columns) of the detector; without correction, an obvious banding pattern would be visible on the image. To correct for differences in gain between the detector elements in the array, a gain image or *flat field image* is acquired. Figure 11-13 shows an image with and without flat field correction. To acquire a flat field image, the array is exposed to a homogeneous signal (a uniform x-ray beam with perhaps a copper filter placed near the x-ray tube to approximate patient attenuation), and a series of images are acquired and averaged. This produces the initial gain image $G'(x,y)$. The dark image, $D(x,y)$, is subtracted from the initial gain image to obtain the gain image: $G(x,y) = G'(x,y) - D(x,y)$. The average gray scale value on the gain image is computed, and this may be designated $\bar{G}$.

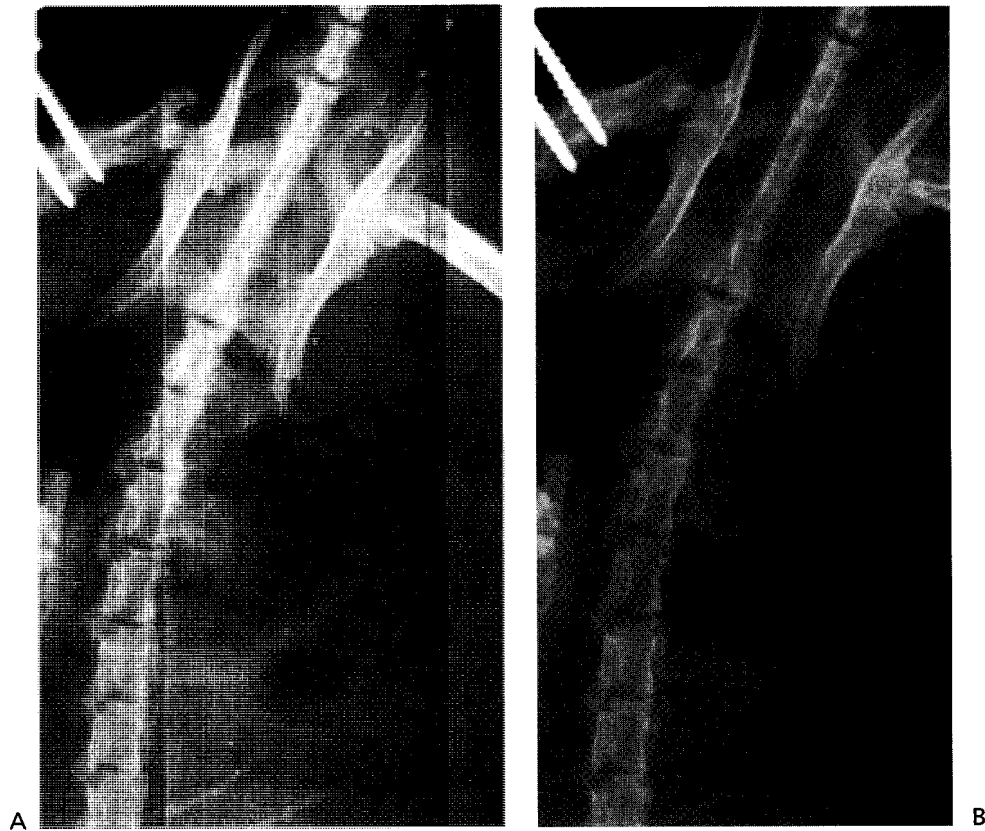


FIGURE 11-13. The raw and corrected images from a prototype flat panel detector array are illustrated. **A:** Obvious banding artifacts are observed on the raw image. **B:** After flat field correction, as described in the text, the artifacts are removed, producing a high quality digital image.

When an initial (raw) medical x-ray image, $I_{raw}(x,y)$, is acquired, it is corrected to yield the final image $I(x,y)$ with the following equation:

$$I(x,y) = \frac{\overline{G}(I_{raw}(x,y) - D(x,y))}{G(x,y)}$$

Virtually all digital x-ray detector systems with discrete detector elements make use of digital image correction techniques. CR systems do not have discrete detectors, and so these techniques are not applied in CR. However, corrections are applied to each row in a CR image to compensate for differences in light conduction efficiency of the light guide. This correction process is conceptually identical to the flat field correction described earlier for discrete detector array systems; however, they are applied in only one dimension.

Global Processing

One of the most common methods of image processing in radiology is performed by altering the relation between digital numbers in the image and displayed bright-

ness. Windowing and leveling of a digital image are common examples of this type of global image processing. Windowing and leveling are simple procedures performed routinely; however, the consequences on the displayed image are profound (Fig. 11-14). Reversing the contrast of an image (changing an image with white bones to one with black bones) is also a global processing technique. Global processing is often performed by the display hardware instead of through manipulation of the image data directly. Mathematically, global processing (including windowing and leveling) is performed by multiplying and adding the gray scale value of each pixel in the image by (the same) constants and then applying thresholding techniques. For example, image $I(x,y)$ can be windowed and leveled to produce image $I'(x,y)$, as follows:

$$I'(x,y) = b \times [I(x,y) - a]$$

Then, limits are placed on image $I'(x,y)$ to constrain the possible gray scale values to within the dynamic range of the display system (usually 256 gray scale units—8 bits). This is done by defining upper (T_{high}) and lower (T_{low}) threshold limits:

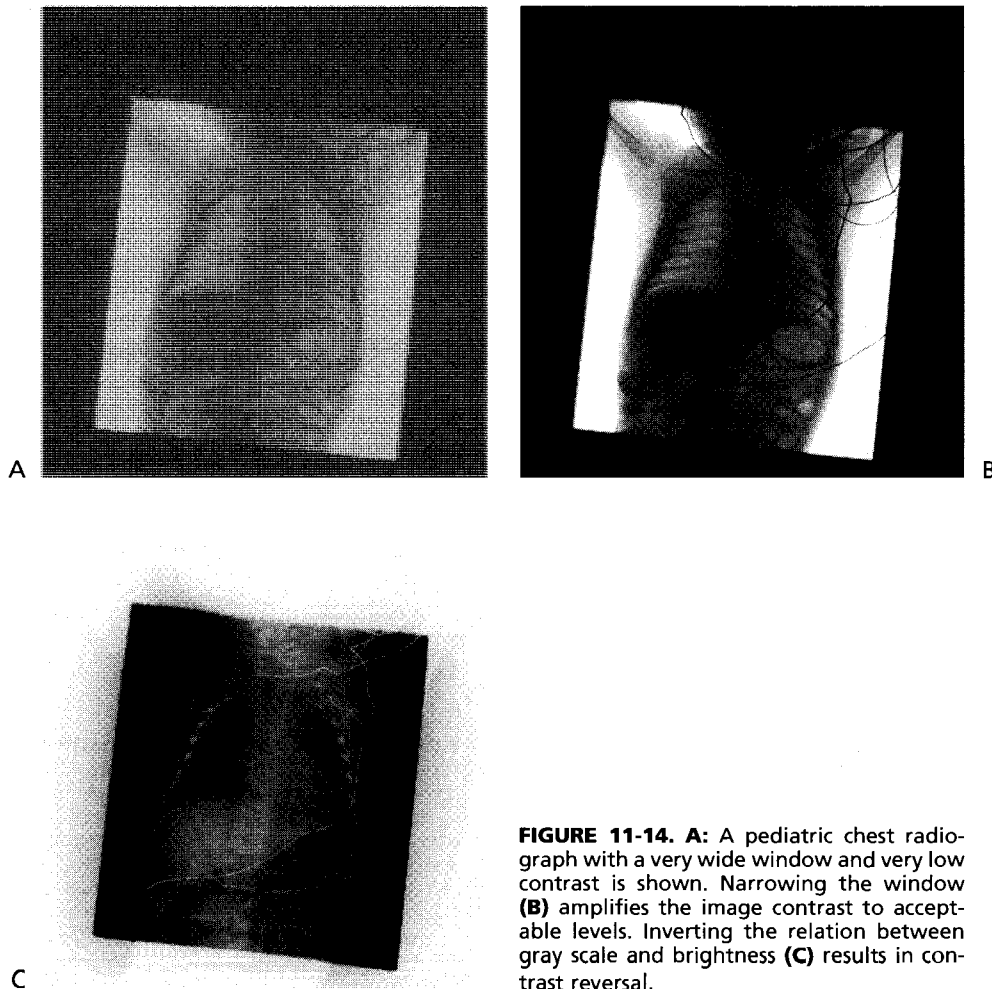


FIGURE 11-14. A: A pediatric chest radiograph with a very wide window and very low contrast is shown. Narrowing the window **(B)** amplifies the image contrast to acceptable levels. Inverting the relation between gray scale and brightness **(C)** results in contrast reversal.

If $I'(x,y) > T_{high}$, then $I'(x,y) = T_{high}$.

If $I'(x,y) < T_{low}$, then $I'(x,y) = T_{low}$.

Notice that the image is inverted in contrast when the constant b is negative.

Image Processing Based on Convolution

The science of manipulating digital images often involves the mathematical operation called *convolution*. Convolution is discussed throughout this text, and especially in the chapters on image quality (Chapter 10) and computed tomography (Chapter 13). Convolution is defined mathematically as follows:

$$g(x) = \int_{-\infty}^{+\infty} I(x')h(x-x')dx'$$

where $I(x)$ is the input, $g(x)$ is the result, and $h(x)$ is called the convolution *kernel*. When applied to digital images, this integral equation results in essentially just shifting and adding gray scale values. Convolution is a straightforward operation that should be understood by all professionals involved in the practice of radiology. Figure 11-15 illustrates convolution as it applies to digital images.

The functional form of the convolution kernel $h(x)$ (i.e., its shape when plotted) can have a profound influence on the appearance of the processed image. Let us take the example of a simple 3×3 convolution kernel, and examine how the nine values that result can be manipulated to achieve different effects. The following ker-

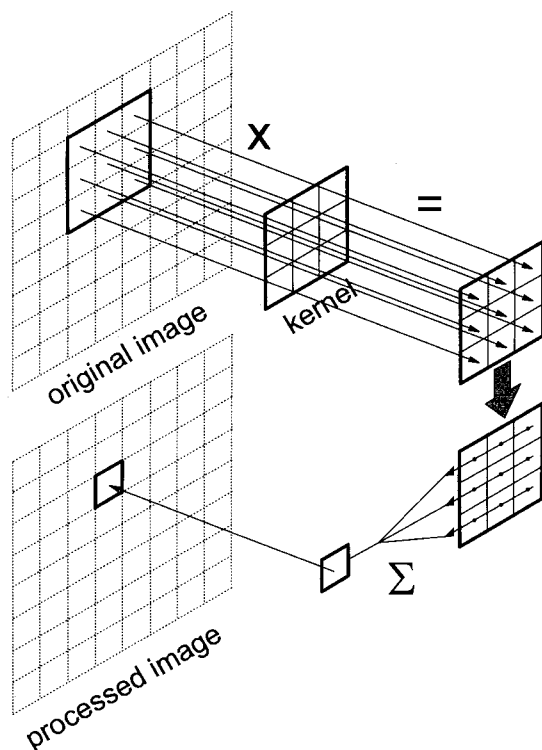


FIGURE 11-15. The concept of discrete convolution as used in digital image processing is illustrated. A 3×3 pixel region of the original image is multiplied by a 3×3 convolution kernel. The nine products of this multiplication are summed, and the result is placed in the pixel in the processed image corresponding to the center of the 3×3 region on the original image. The convolution operation moves across the entire image, pixel by pixel.

nel is called a *delta function*, and convolving with it does not change an image in any way:

0	0	0
0	1	0
0	0	0

An entire image can be shifted one pixel to the left by the following kernel:

0	0	0
1	0	0
0	0	0

When all the values in the kernel are positive and non-zero, the kernel blurs an image. For example, the following kernel smooths some of the noise in an image but also reduces its resolution somewhat:

1	1	1
1	1	1
1	1	1

Smoothing with a 3×3 kernel produces a subtle blurring effect that is not well appreciated on an image at textbook resolution. However, the convolution can be performed using a larger kernel (e.g., 11×11) to achieve substantial blurring effects that are easy to see (Fig. 11-16B).

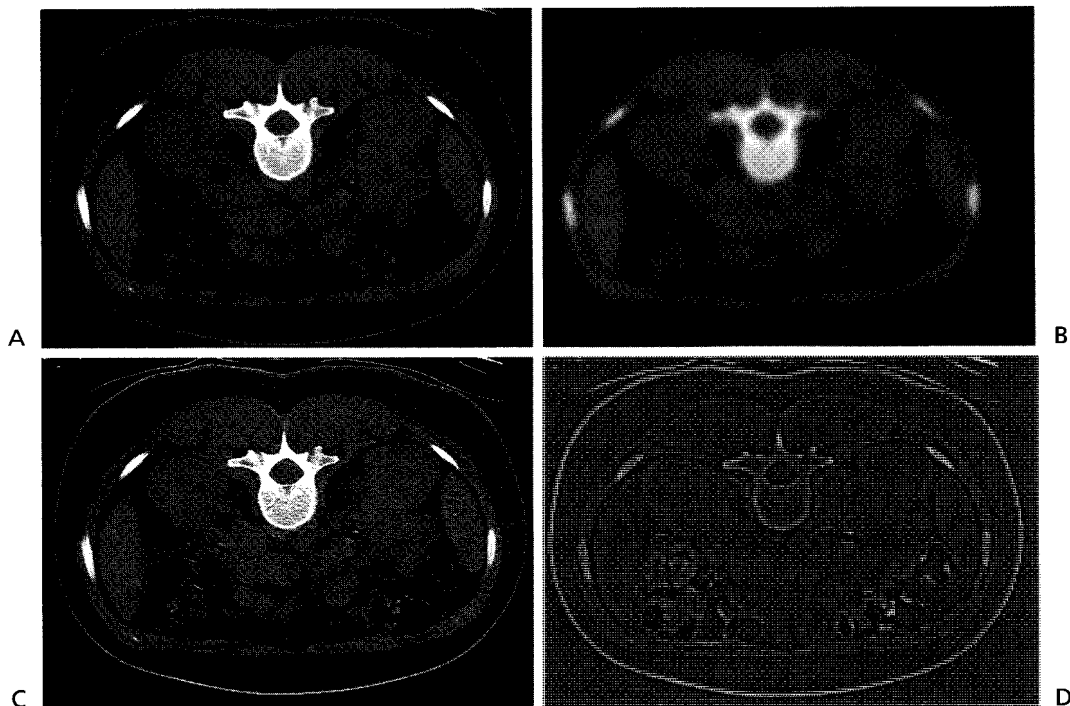


FIGURE 11-16. The images shown illustrate the effects of various image-processing procedures. (See text for details.) **A:** The original image. **B:** This image was blurred by application of an 11×11 convolution kernel. **C:** An edge-enhanced image is shown. **D:** The effect of harmonization is shown.

When the 3×3 kernel containing all ones is convolved with an image, the sum of all the gray scale values in the resulting image is nine times greater than the sum of all the gray scale values in the original image. To correct for this, the resulting image is divided by 9 (pixel by pixel). Another way to perform this normalization is to divide all the kernel values by 9 before convolution is performed:

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

When the kernel contains both positive and negative numbers, the image, instead of being smoothed, will be sharpened. For example, the influence of the following kernel is illustrated in Fig. 11.16C.

-1	-1	-1
-1	9	-1
-1	-1	-1

This edge-sharpening convolution filter is conceptually similar to the types of kernels used in CT convolution backprojection. The edge-sharpening characteristics of a CT convolution kernel are designed to compensate for the blurring effects of backprojection.

An image that has been smoothed with the use of a blurring kernel can be subtracted from the original image to produce a *harmonized* image. The original image contains both low- and high-frequency components, but in the blurred image high-frequency structures are dampened. The difference image tends to reduce low-frequency components in the image, accentuating the high-frequency structures (edges), as shown in Fig. 11-16D.

Convolution in principle can be undone by deconvolution; however, a noise penalty often occurs. Deconvolution is essentially the same thing as convolution, except that the shape of the deconvolution kernel is designed to reverse the effects of a convolution kernel or physical process. For example, if an image is convolved with a kernel K , and then is convolved with its inverse K^{-1} (i.e., deconvolved), the original image will be reproduced.

Median Filtering

Convolution with a 3×3 kernel of all ones is a process of averaging nine adjacent pixels and storing the average value in the center pixel of the new image. The average is affected by outlying values (high or low), whereas the median is the middle value and is unaffected by outliers.

Multiresolution or Multiscale Processing

An entity that exhibits similar properties (spatial or otherwise) over a broad range of scale factors is considered to exhibit *fractal* behavior. The popularity of fractals in the past decade has led to a class of image-processing algorithms that can be called multiscale. Take as an example an input image having 512×512 pixels. This original image is broken down into a set of images with different matrix sizes: 256×256 , 128×128 , 64×64 , 32×32 , 16×16 , and so on. Each of the smaller matrix images is computed by averaging four pixels into one; for example, adjacent 2×2

pixels are averaged to go from a 512×512 image to a 256×256 image, and then to a 128×128 image, and so on. Image-processing techniques are then applied to each of the smaller-resolution images, and they are then recombined to get a single, processed 512×512 image.

Adaptive Histogram Equalization

Medical images sometimes exhibit a very large range of contrast; for example, the average CT numbers in the lung fields are markedly different than in the liver, and these are different from the CT numbers of bone. With such a broad range of gray scale, the radiologist often reviews CT images with two or three different window/level settings, because one window/level setting is insufficient to visualize all parts of the image. This problem can be rectified in some types of images by adaptive histogram equalization (AHE). Essentially, AHE algorithms apply different window/level settings to different parts of the image, with a smooth spatial transition. AHE images have an appearance similar to the image shown in Fig. 11-16D.

11.10 CONTRAST VERSUS SPATIAL RESOLUTION IN DIGITAL IMAGING

When an analog image is partitioned into a digital matrix of numbers, compromises are made. As an example, screen-film mammography produces images with measurable spatial resolution beyond 20 cycles/mm. To obtain this resolution with a digital image, 25- μm pixels would be required, resulting in a $7,200 \times 9,600$ pixel image. Images of this size cannot be displayed (at full resolution) on any monitor, and storing huge images is expensive as well. Consequently, digital radiographic images are invariably lower in spatial resolution than their analog screen-film counterparts (Fig. 11-17), both for mammography and for general diagnostic radiography.

Images in digital format have advantages as well. The ability to transmit images electronically, to produce identical copies that can be in multiple locations at the

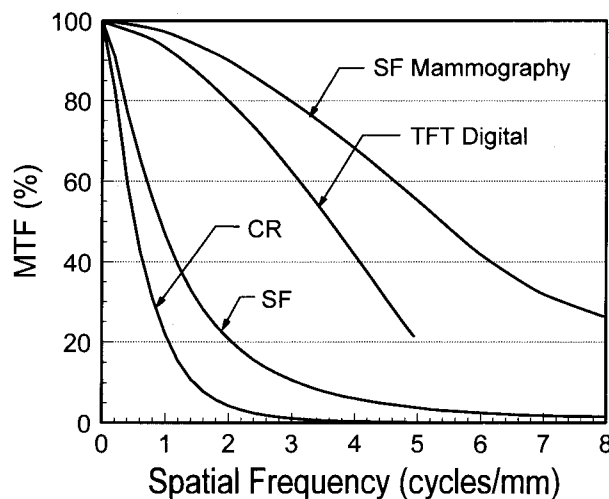


FIGURE 11-17. The modulation transfer functions (MTFs) are shown for several modalities. SF, screen-film; CR, computed radiography; SF Mammography, screen-film mammography; TFT Digital, direct detection 100- μm pixel flat panel detector system.

same time, and to archive images using computers instead of long rows of shelves represent practical advantages. Digital technology is clearly where diagnostic radiology is headed. In addition to mere workflow issues, however, digital images offer a number of advantages with respect to improving diagnoses. The ability to readjust the contrast of an image after detection is an underappreciated feature of digital images that leads to improved interpretation and better diagnosis. Digital detector systems may not produce spatial resolution equal to that of film, but the contrast resolution (at equivalent radiation dose levels) is superior for most digital radiographic systems compared with screen-film systems. The combination of improved contrast resolution (better DQE, discussed in Chapter 10) and improved contrast (window/level) can lead to an overall improvement in diagnostic performance for a significant fraction of clinical examinations. Under a wide array of circumstances, the improvement in contrast resolution that digital images provide outweighs the slight loss in spatial resolution.

ADJUNCTS TO RADIOLOGY

12.1 GEOMETRIC TOMOGRAPHY

Geometric tomography, also called body tomography or conventional tomography, makes use of geometric focusing techniques to achieve the tomographic effect. A planar image receptor, such as a screen-film cassette or a computed radiography cassette, is used. The principles of geometric tomography are illustrated in Fig. 12-1. The patient is positioned on the table, and the desired plane in the patient to be imaged is placed at the pivot point of the machine. The location of the pivot point with respect to the patient defines the focal plane (i.e., the tomographic plane to be imaged). During the x-ray exposure, the x-ray tube travels through an arc above the patient. The imaging cassette travels in the opposite direction, so that the x-rays remain incident upon it during the entire exposure. The angle through which the x-ray tube travels while the x-rays are on is called the *tomographic angle*. With a larger tomographic angle, the anatomy away from the focal plane is blurred more effectively, and the effective slice thickness decreases.

Geometric tomography achieves the tomographic effect by blurring out the x-ray shadows of objects that are above and below the focal plane. Figure 12-2A illustrates this concept. The circle is above the focal plane, the square is at the focal plane, and the triangle is below the focal plane. Imagine that these simple objects represent the patient's anatomy at those three positions. During the acquisition, the arc motion of the x-ray tube causes the x-ray shadows of out-of-plane objects (i.e., the circle and the triangle) to be blurred across the image receptor. Meanwhile, the x-ray shadows of objects that are in the focal plane (the square) remain aligned on the cassette during the entire exposure. Thus, the contrast of in-plane objects accumulates at one place on the image receptor, while the contrast of out-of-plane objects is spread out over the receptor. Spreading the contrast of an object out over a large area on the image causes it to be blurred out and therefore less visible to the eye.

Figure 12-2B illustrates the location of the three objects at the beginning of the exposure (A) and at the end (B). The shadow of the triangle begins on the left side of the image (at position A) and is blurred to the right (at position B). The upper arrow illustrates the blurred path of the triangle's shadow during the exposure. The shadow of the circle is on the right side of the image receptor at the beginning of the exposure (position A), and it ends up on the left side of the cassette at the end of the exposure (position B). The lower arrow on Fig. 12-2B illustrates the path of the x-ray shadow of the circle during the exposure. The shadow of the square was

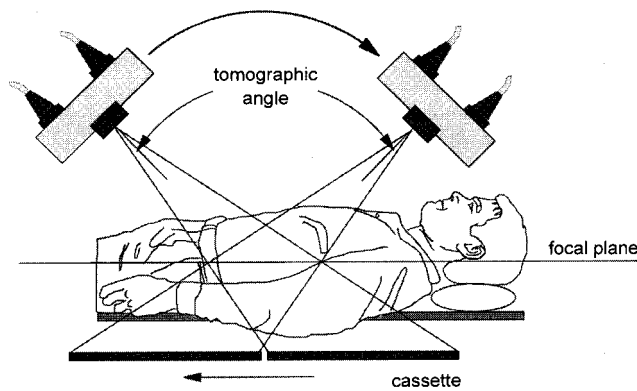


FIGURE 12-1. The motion of the x-ray tube and cassette that occurs in geometric tomography is illustrated. The focal plane is the location of the tomographic slice, and this occurs at the pivot point of the machine.

aligned at the same position on the detector throughout the entire exposure—therefore, the contrast of the square was summed up at the same point on the image and it is highly visible.

The use of geometric tomography has gradually waned over the years since computed tomography was introduced. However, geometric tomography is a less expensive procedure, and it is still used extensively for intravenous pyelograms (IVPs), and to a lesser extent for some temporal bone and chest radiographic applications. The out-of-plane anatomy of the patient is not removed from the image, as it is in computed tomography; it is merely blurred out over the image. The blurred out back-

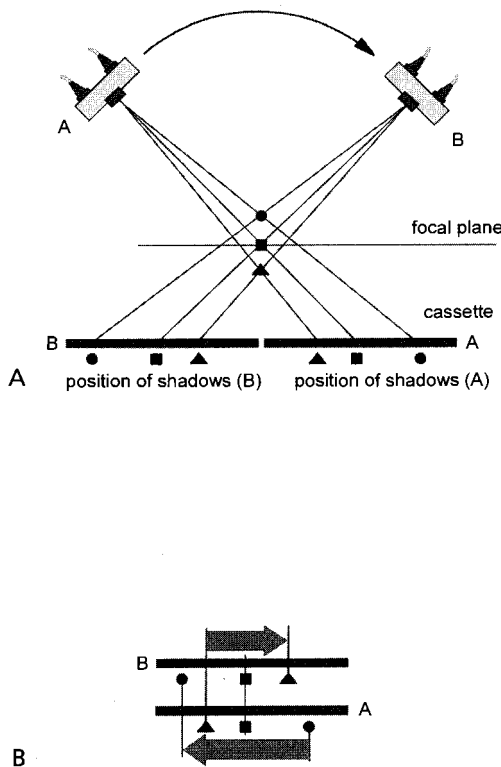


FIGURE 12-2. A: This figure illustrates how geometric tomography works. At the beginning of the exposure (x-ray tube and cassette in position A) the three objects (circle, square, and triangle) are projected as illustrated. At the end of the exposure (position B), the x-ray shadows of the objects are in different locations on the image. As the x-ray tube moves through the tomographic angle, the x-ray shadows of objects above the focal plane are blurred from right to left across the cassette, while those of objects below the focal plane are blurred in the other direction. **B:** The blur distances of the circle, square, and triangle are shown. The top arrow indicates the path through which the triangle's shadow was blurred. The bottom arrow shows the blur path of the circle. The square was in the focal plane, so its shadow did not move during the scan and therefore the square is in focus.

ground anatomy causes geometric tomography images to be inherently low in contrast, so this modality is used only for very high-contrast imaging situations—for example, when contrast agent is present in the kidney (IVPs), for air-bone contrast of the inner ear, or for imaging tissue-air contrast in the lungs. In the past, specialized radiographic units (dedicated tomography systems) were produced by several manufacturers and were capable of a variety of motions. In addition to the linear motion that is commonly available on standard radiographic rooms (as an option) today, dedicated tomographic systems had circular, tri-spiral, and other complicated motions. These nonlinear motions produced blurring patterns that were thought to be more useful in some circumstances. However, dedicated tomography systems have limited use in today's radiology departments and are no longer sold commercially.

The notion of slice thickness in geometric tomography is not as tangible as it is in computed tomography. The object blurring that was illustrated in Fig. 12-2 is really a continuous effect—there is only a minute amount of blurring of objects that are located precisely at the focal plane (related to the size of the object), and the amount of blur for out-of-plane objects increases as the distance from the object to the focal plane increases. For larger tomographic angles, the blurring that an out-of-plane object experiences is greater than for smaller tomographic angles. Figure 12-3 illustrates the slice sensitivity profiles for two different tomographic angles. Notice that the amplitude of the slice sensitivity profile for objects even 20 cm away from the focal plane is still non-zero, meaning that ghost shadows from such objects remain in the image. By comparison, the slice sensitivity profiles in computed tomography become zero just outside the collimated slice thickness.

Geometric tomography requires an x-ray exposure of the entire thickness of the patient to acquire just one tomographic image. If a series of tomographic images at different focal depths are to be acquired, the tissue volume in the patient is exposed repeatedly for each acquisition. If the dose from one tomographic acquisition is 5 mGy (0.5 rad) and ten tomographic slices are acquired, then the dose to the patient increases to 50 mGy (5 rad). By comparison, if the dose from one computed tomographic slice is 30 mGy (3 rad), then the dose from ten contiguous slices is still 30 mGy (3 rad), excluding the dose resulting from scattered radiation. However, a

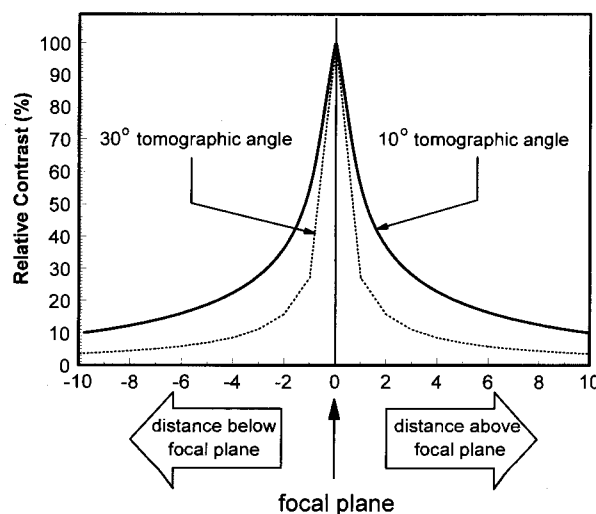


FIGURE 12-3. The slice sensitivity profiles for geometric tomography are illustrated. Thinner slices (sharper peaked profiles) occur when the tomographic angle is greatest. Tomograms with a 10-degree tomographic angle result in very thick slices and are called zonograms. Notice that for geometric tomography, the slice sensitivity profiles never go completely to zero, unlike those that are characteristic in computed tomography.

larger tissue volume is exposed. This implies that geometric tomography is a reasonable and cost-effective approach for imaging situations requiring a single slice (e.g., IVP) or for a few slices, but when a lot of tomographic slices are required, geometric tomography becomes a very high-dose procedure and computed tomographic techniques should be considered instead (with multiplanar reconstruction if coronal or sagittal slice planes are desired).

12.2 DIGITAL TOMOSYNTHESIS

Digital tomosynthesis is almost completely analogous to geometric tomography. However, the technique requires a modest amount of computer processing, so the images acquired must be digital. Practical applications of digital tomosynthesis make use of digital detectors such as flat panel imaging devices, which are capable

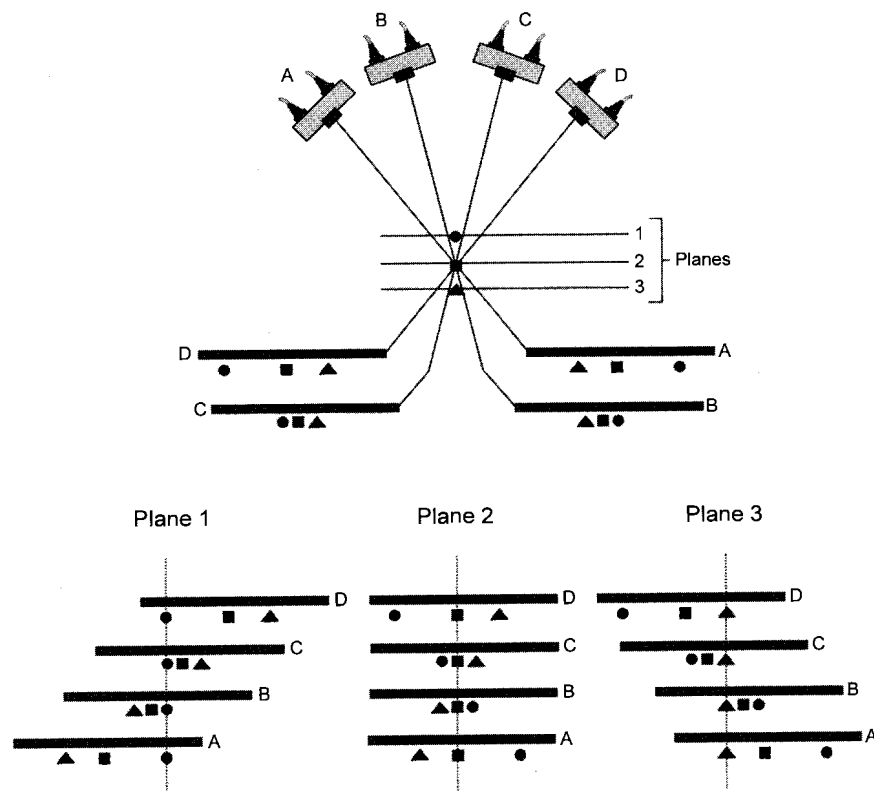


FIGURE 12-4. The principles of tomosynthesis are demonstrated. Four digital images (A, B, C, and D) are acquired at different angles, as shown in the top of the figure. Usually about ten images are acquired, but four are sufficient to illustrate the concept. The digital images are shifted and then summed to produce tomographic images. Different shift distances can be used to reconstruct different tomographic planes in the volume. For example, with no shift distance, the images are aligned and plane 2 is in focus, which is coincident with the focal plane of the scan: plane 2 is just a digital example of regular geometric tomography. Plane 1 is focused using the shift distances illustrated at the bottom left of the figure, and plane 3 can be focused using the shifts illustrated at the bottom right. After shifting the images in the computer, they are added to produce the tomographic image.

of relatively rapid readout, allowing the scan to be performed in a short period of time. In tomosynthesis, instead of blurring the image data over the arc of the tube rotation onto a single image, a series of perhaps ten different images is acquired at ten different tube angles. Figure 12-4 illustrates this concept using four different images. The very important property that distinguishes tomosynthesis from geometric tomography is that, with tomosynthesis, any plane can be reconstructed with just one acquisition. Therefore, one set of images acquired over an arc, as shown in Fig. 12-4, can be used to reconstruct slices at different planes in the patient.

The bottom panels in Fig. 12-4 demonstrate how these reconstructions are accomplished. To reconstruct plane 1, in which the circle resides, the images are shifted as illustrated on the lower left panel and the image data are added. Notice how the shadows of the circle are superimposed because of the shifting of the image data. The resulting image will appear (in principle) identical to a geometric tomography image focused at plane 1. Reconstruction of other planes, for example plane 3, is accomplished by simply shifting the individual image data by different amounts before summing. In the bottom right panel, the image shifts were adjusted so that the triangles were aligned, and in fact all objects that are in plane 3 are in focus. Using this approach, all planes of interest in the acquisition volume can be reconstructed.

Tomosynthesis represents a much more dose-efficient approach to geometric tomography than older single-film techniques. The development of flat panel digital detectors was an enabling technology for digital tomosynthesis: Image intensifiers with TV cameras and video digitizers are capable of digital image acquisition, but the markedly curved surface of the image intensifier precludes the use of simple and practical reconstruction methods. In addition to the shift-and-sum procedure used to reconstruct the images as illustrated in Fig. 12-4, other, more sophisticated mathematical deblurring techniques can be used to improve the image quality and slice sensitivity profile.

12.3 TEMPORAL SUBTRACTION

Digital Subtraction Angiography

The most common example of temporal subtraction is digital subtraction angiography (DSA). In DSA, a digital radiographic image of the patient's anatomy (the "mask") is acquired just before the injection of a radiopaque contrast agent. A sequence of images is acquired during and after the contrast agent injection. The mask image is subtracted from the images containing contrast agent. When the patient does not move during the procedure, the subtraction images provide exquisite detail of the vascular anatomy and the other anatomic aspects of the patient (e.g., bones) are subtracted out of the image. Because the background anatomy is eliminated, the subtracted images can be processed (windowed and leveled), amplifying the contrast in the vessels. If some of the patient's anatomic landmarks are needed for orientation, the enhanced vascular image can be overlaid on a faint version of the mask image, mathematically weighted to reduce contrast. DSA has in general replaced older film subtraction techniques.

Before the subtraction of images, the raw image data are logarithmically processed. If the thickness of the background tissue is t_{bg} with a linear attenuation coefficient of μ_{bg} , and the vessel thickness is t_{vessel} with a linear attenuation coefficient

cient of μ_{vessel} , the mask image (I_m) and the contrast image (I_c) are given by the following equations:

$$I_m = N_o e^{-\mu_{bg} t_{bg}}$$

$$I_c = N_o e^{-\mu_{vessel} t_{vessel} - \mu_{bg} t_{bg}}$$

where N_o is the incident number of x-rays. Taking the logarithm of the two equations and then subtracting yields the subtracted image I_s :

$$I_s = \ln(I_m) - \ln(I_c) = \mu_{vessel} t_{vessel}$$

Of course the actual equations include integration over the x-ray spectrum, but the simplified equations given here illustrate the point that the subtracted signal is linear with vessel thickness, as long as the iodine concentration (i.e., μ_{vessel}) is constant over the region of interest in the vessel. In principle, on a well-calibrated DSA

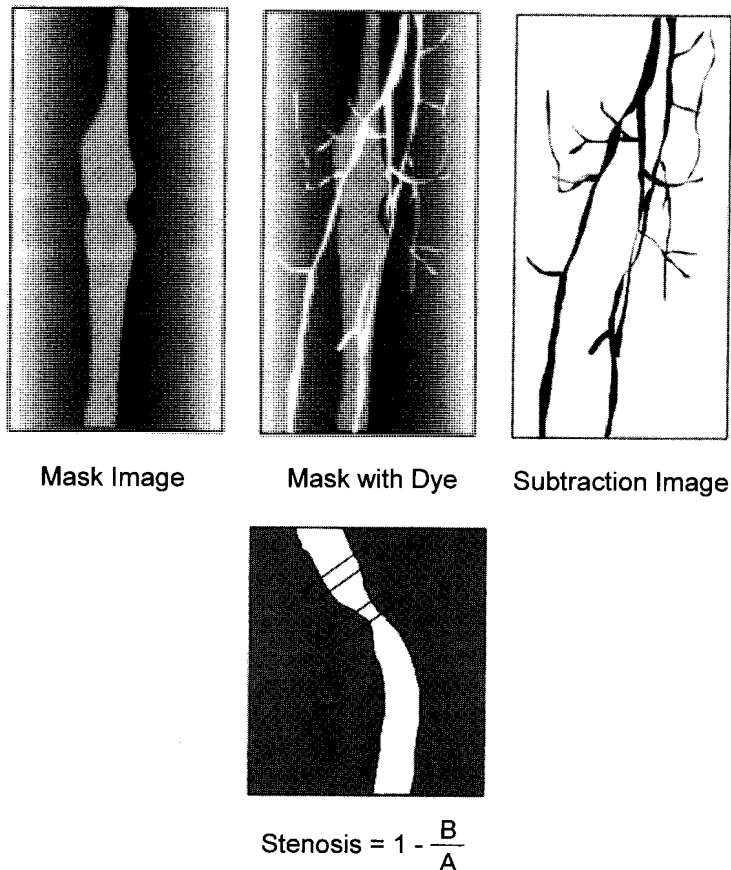


FIGURE 12-5. In digital subtraction angiography (DSA), a mask image containing only patient anatomy is acquired before dye injection, and then a series of images are acquired as the iodine bolus passes through the patient's vasculature. Subtraction of the dye images from the mask results in the DSA subtraction image, which demonstrates only the vascular anatomy (i.e., the patient's background anatomy is subtracted out). DSA is in principle a quantitative technique, and vascular stenosis (based on vessel areas) can be computed as illustrated.

system, the amount of vascular stenosis can be quantified by summing the gray scale values in two bands, one placed across the stenosis and the other just adjacent to it (Fig. 12-5). This is called video densitometry.

Other Temporal Subtraction Systems

Temporal subtraction works best when the time interval between the two subtracted images is short (seconds)—digital subtraction angiography is a good example. However, radiologists commonly make comparisons between images that are acquired over longer time intervals in order to assess serial changes of a condition or disease status in the patient. A typical clinical example is in chest radiography, where two images obtained 1 year apart may be compared. If the images are digital, they can be subtracted; this would generally make any changes more obvious than if each image were examined separately. However, positioning and body habitus differences between images acquired years apart usually result in poor subtraction due to misregistration. To mitigate this problem, software that uses the patient's own hard anatomy (e.g., rib crossings, vertebral bodies, scapula) as fiducial references allows one image to be nonlinearly warped and spatially aligned with the other one. Once the images are aligned and subtracted, subtler features such as pulmonary nodules may be identified with higher conspicuity. Although this technique is still in the research phase, given the growing use of computers in the display and analysis of medical images, it is likely to become a part of the standard repertoire of image analysis tools available for diagnostic imaging.

12.4 DUAL-ENERGY SUBTRACTION

Dual-energy subtraction is used to exploit differences between the atomic numbers of bone and soft tissue. The atomic number (Z) of calcium is 20, and the effective atomic number of bone (a mixture of calcium and other elements) is about 13. The effective Z of tissue (which is mostly hydrogen, carbon, nitrogen, and oxygen) is about 7.6. The attenuation of low-energy x-rays (e.g., a 50 kVp x-ray spectrum) is very dependent on Z —a result of the Z^3 dependence of the photoelectric effect, which is the dominant type of x-ray interaction at lower x-ray energies. However, at higher energies (e.g., a copper-filtered 120-kVp x-ray spectrum), the Compton interaction dominates, and such images are less dependent on Z and more on physical density. The bottom line is that the contrast on a standard digital radiographic image changes with the x-ray energy, and the amount of the change differs depending on the atomic number of the structure.

Dual-energy subtraction starts with the acquisition of two digital projection radiographic images, one at a low energy and the other at a high energy. This can be done simultaneously using a sandwiched detector, or it can be done using two x-ray exposures at different kVp (Fig. 12-6). The mathematics is simplified here, but the two images (I_1 and I_2) are subjected to weighted logarithmic subtraction, essentially $\ln(I_1) - R \ln(I_2)$, where R is a constant that is used to change the weighting. By adjustment of R , the soft tissue parts in the image can be accentuated or the bones in the image can be amplified. In principle, the contrast from either bone or soft tissue can be completely subtracted out, but pragmatic realities such as beam hardening, scattered radiation, and quantum noise reduce this ideal. Nevertheless,

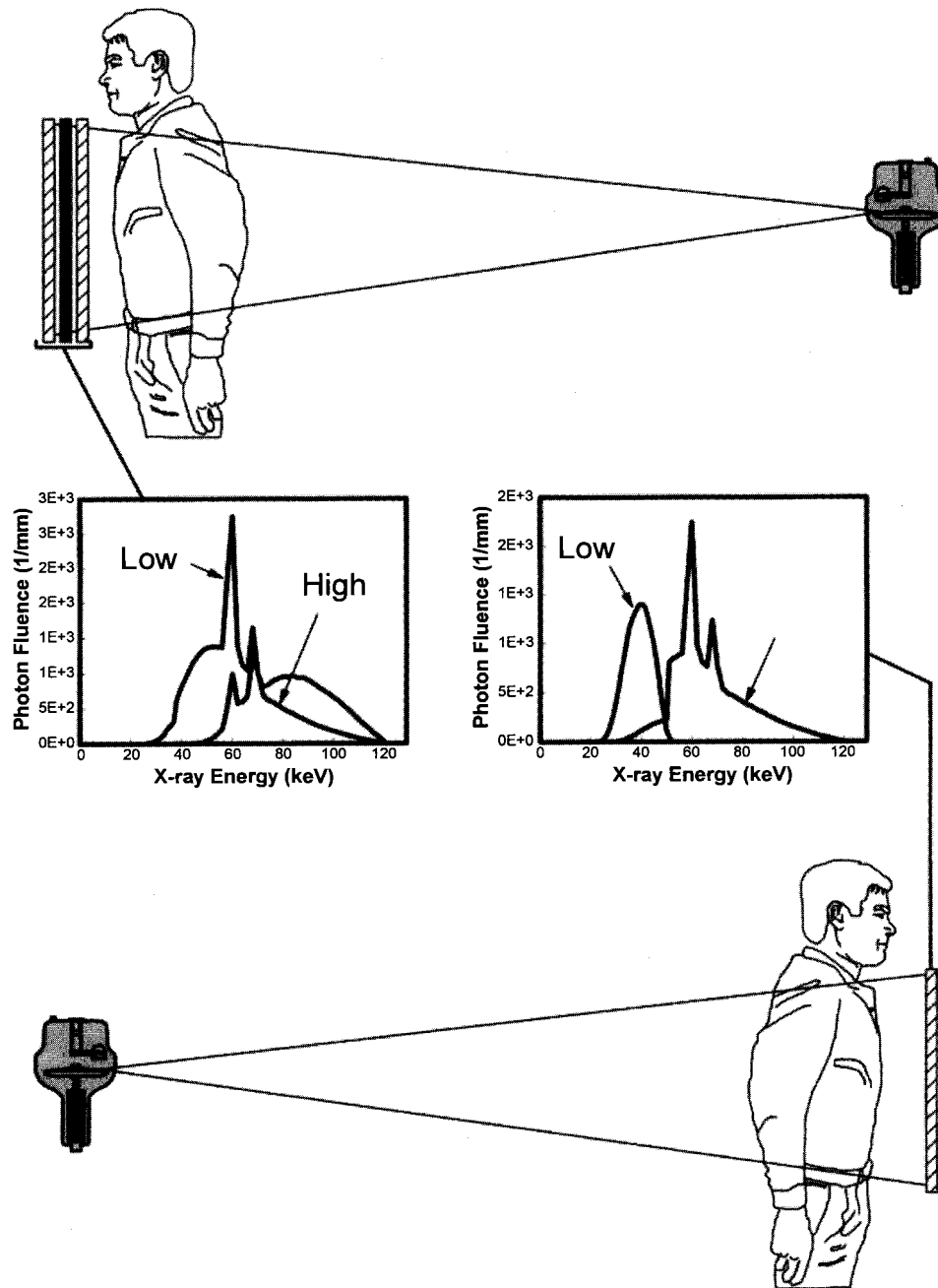


FIGURE 12-6. Two types of acquisition for dual-energy radiography are shown. **Top:** A sandwiched detector (detector 1, metallic filter, detector 2) is used with a single, high-kVp x-ray spectrum for the single-shot dual-detector approach. The front detector absorbs more of the low-energy x-rays in the beam, and the back detector absorbs more of the higher-energy x-rays, because the x-ray beam is hardened by the front detector and by the filter placed between the two detectors. The low- and the high-energy absorbed spectra are illustrated (120 kVp, 100-mg/cm² BaFBr screens were used as detectors, and two BaFBr screens were used as the mid-detector filter). **Bottom:** A two-pulse single-detector approach to dual-energy imaging. One detector detects two x-ray pulses (of high and low kVp), which are acquired over a short time interval. The resulting absorbed x-ray spectra are shown (50-kVp, 1-mm filtration and 120-kVp, 9-mm Al filtration with one 60-mg/cm² Gd₂O₂S detector).

dual-energy subtraction is available clinically on some systems for chest radiography. In chest radiography, the pulmonary structures are more of a focus than the ribs, and it would be advantageous to eliminate contrast from the ribs so that they do not obscure subtle details in the pulmonary area of the image. Figure 12-7 illustrates the low-energy, high-energy, bone-only, and soft tissue-only images obtained using dual-energy radiography.

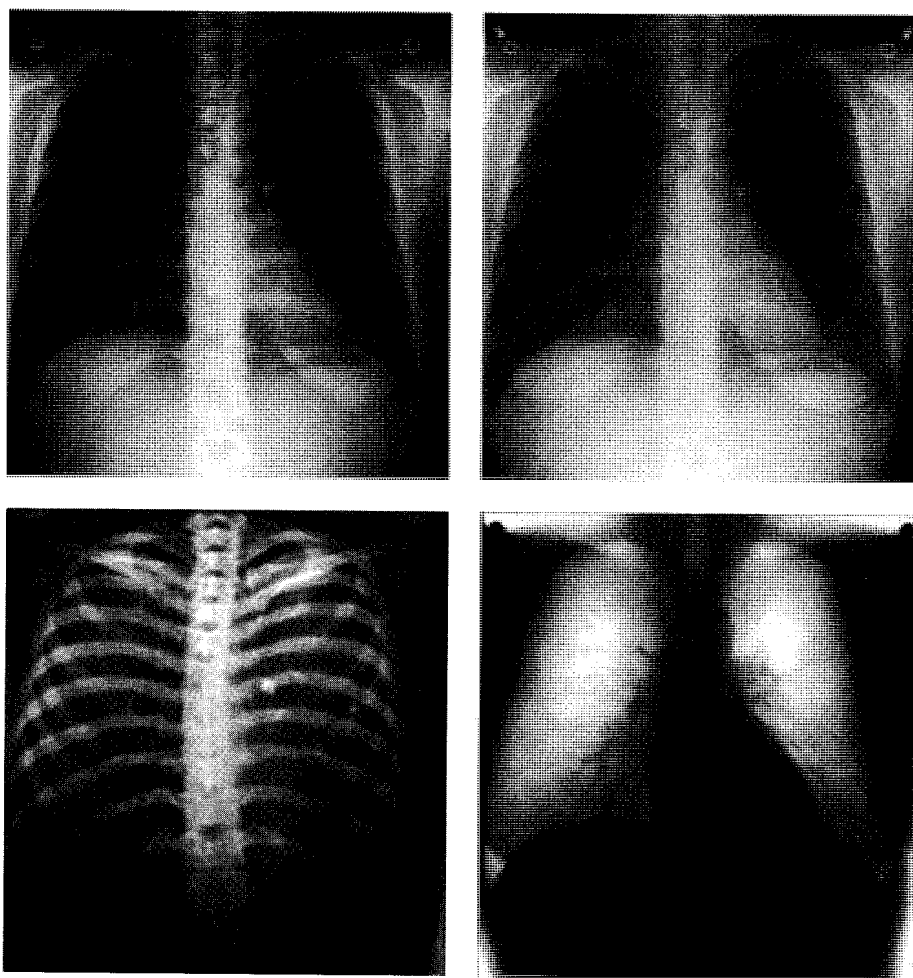


FIGURE 12-7. Dual-energy images are shown. The upper left image is the low-energy image (56 kVp); the upper-right image is the high-energy image (120 kVp, 1 mm Cu). The lower left image shows the energy subtraction image weighted to present bone only, and on the lower right is the tissue-weighted image. This patient had a calcified granuloma that is well seen on the bone-only image. The soft tissue image shows the lung parenchyma without the overlaying ribs as a source of distraction.

COMPUTED TOMOGRAPHY

Computed tomography (CT) is in its fourth decade of clinical use and has proved invaluable as a diagnostic tool for many clinical applications, from cancer diagnosis to trauma to osteoporosis screening. CT was the first imaging modality that made it possible to probe the inner depths of the body, slice by slice. Since 1972, when the first head CT scanner was introduced, CT has matured greatly and gained technological sophistication. Concomitant changes have occurred in the quality of CT images. The first CT scanner, an EMI Mark 1, produced images with 80×80 pixel resolution (3-mm pixels), and each pair of slices required approximately 4.5 minutes of scan time and 1.5 minutes of reconstruction time. Because of the long acquisition times required for the early scanners and the constraints of cardiac and respiratory motion, it was originally thought that CT would be practical only for head scans.

CT is one of the many technologies that was made possible by the invention of the computer. The clinical potential of CT became obvious during its early clinical use, and the excitement forever solidified the role of computers in medical imaging. Recent advances in acquisition geometry, detector technology, multiple detector arrays, and x-ray tube design have led to scan times now measured in fractions of a second. Modern computers deliver computational power that allows reconstruction of the image data essentially in real time.

The invention of the CT scanner earned Godfrey Hounsfield of Britain and Allan Cormack of the United States the Nobel Prize for Medicine in 1979. CT scanner technology today is used not only in medicine but in many other industrial applications, such as nondestructive testing and soil core analysis.

13.1 BASIC PRINCIPLES

The mathematical principles of CT were first developed by Radon in 1917. Radon's treatise proved that an image of an unknown object could be produced if one had an infinite number of projections through the object. Although the mathematical details are beyond the scope of this text, we can understand the basic idea behind tomographic imaging with an example taken from radiography.

With plain film imaging, the three-dimensional (3D) anatomy of the patient is reduced to a two-dimensional (2D) projection image. The density at a given point on an image represents the x-ray attenuation properties within the patient along a line between the x-ray focal spot and the point on the detector corresponding to the

point on the image. Consequently, with a conventional radiograph of the patient's anatomy, information with respect to the dimension parallel to the x-ray beam is lost. This limitation can be overcome, at least for obvious structures, by acquiring both a posteroanterior (PA) projection and a lateral projection of the patient. For example, the PA chest image yields information concerning height and width, integrated along the depth of the patient, and the lateral projection provides information about the height and depth of the patient, integrated over the width dimension (Fig. 13-1). For objects that can be identified in both images, such as a pulmonary nodule on PA and lateral chest radiographs, the two films provide valuable location information. For more complex or subtle pathology, however, the two projections are not sufficient. Imagine that instead of just two projections, a series of 360 radiographs were acquired at 1-degree angular intervals around the patient's

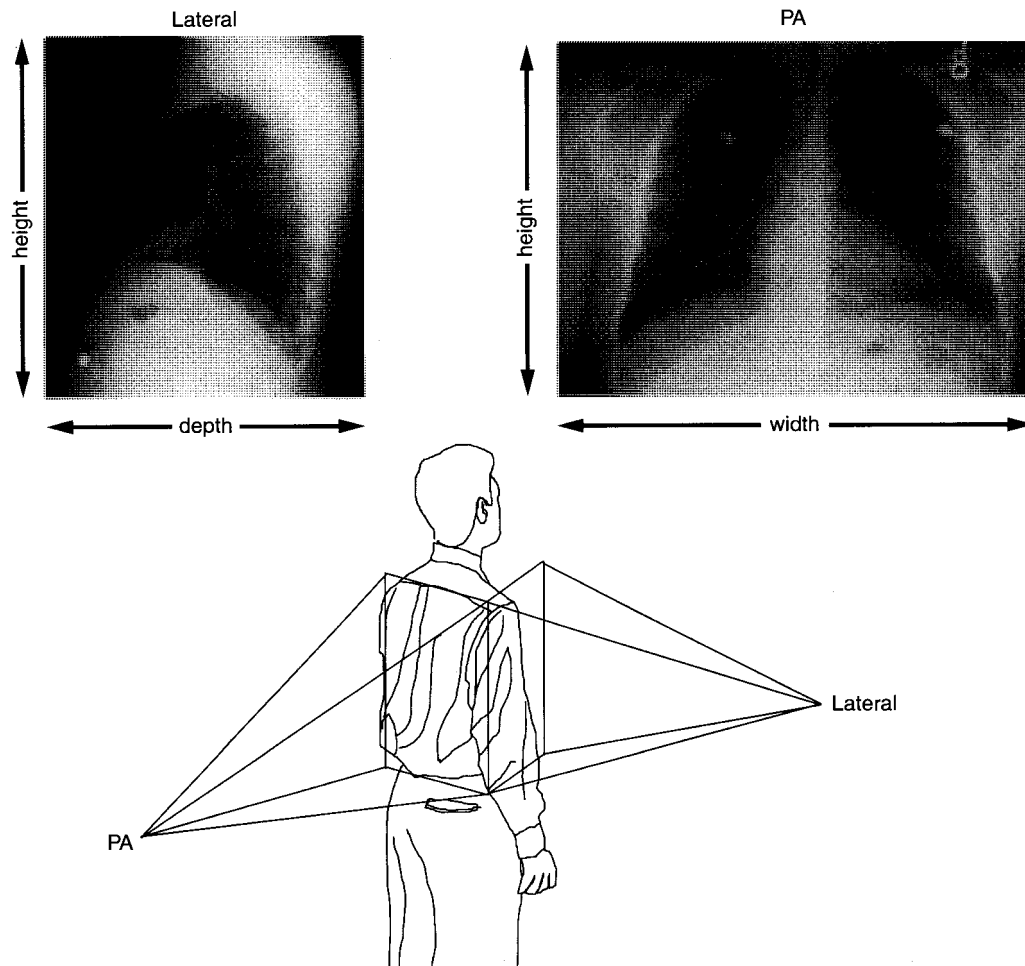


FIGURE 13-1. Posteroanterior and lateral chest radiographs give three-dimensional information concerning the location of an abnormality. In computed tomography, the two views shown here are extended to almost 1,000 views, and with appropriate computer processing true three-dimensional information concerning the patient's anatomy can be displayed.

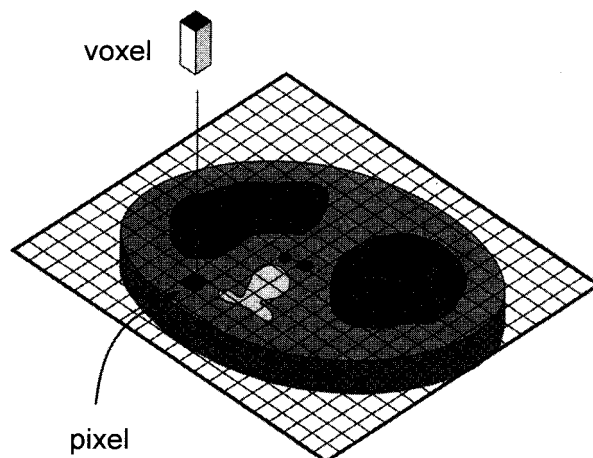


FIGURE 13-2. A pixel (picture element) is the basic two-dimensional element of a digital image. Computed tomographic (CT) images are typically square arrays containing 512×512 pixels, each pixel representing 4,096 possible shades of gray (12 bits). Each pixel in the CT image corresponds to a voxel (volume element) in the patient. The voxel has two dimensions equal to the pixel in the plane of the image, and the third dimension represents the slice thickness of the CT scan.

thoracic cavity. Such a set of images provides essentially the same data as a thoracic CT scan. However, the 360 radiographic images display the anatomic information in a way that would be impossible for a human to visualize: cross-sectional images. If these 360 images were stored into a computer, the computer could in principle reformat the data and generate a complete thoracic CT examination.

The tomographic image is a picture of a slab of the patient's anatomy. The 2D CT image corresponds to a 3D section of the patient, so that even with CT, three dimensions are compressed into two. However, unlike the case with plain film imaging, the CT slice-thickness is very thin (1 to 10 mm) and is approximately uniform. The 2D array of pixels (short for *picture elements*) in the CT image corresponds to an equal number of 3D voxels (*volume elements*) in the patient. Voxels have the same in-plane dimensions as pixels, but they also include the slice thickness dimension. Each pixel on the CT image displays the average x-ray attenuation properties of the tissue in the corresponding voxel (Fig. 13-2).

Tomographic Acquisition

A single transmission measurement through the patient made by a single detector at a given moment in time is called a *ray*. A series of rays that pass through the patient at the same orientation is called a *projection* or *view*. There are two projection geometries that have been used in CT imaging (Fig. 13-3). The more basic type is *parallel beam geometry*, in which all of the rays in a projection are parallel to each other. In *fan beam geometry*, the rays at a given projection angle diverge and have the appearance of a fan. All modern CT scanners incorporate fan beam geometry in the acquisition and reconstruction process. The purpose of the CT scanner hardware is to acquire a large number of transmission measurements through the patient at different positions. The acquisition of a single axial CT image may involve approximately 800 rays taken at 1,000 different projection angles, for a total of approximately 800,000 transmission measurements. Before the axial acquisition of the next slice, the table that the patient is lying on is moved slightly in the cranial-caudal direction (the "z-axis" of the scanner), which positions a different slice of tissue in the path of the x-ray beam for the acquisition of the next image.

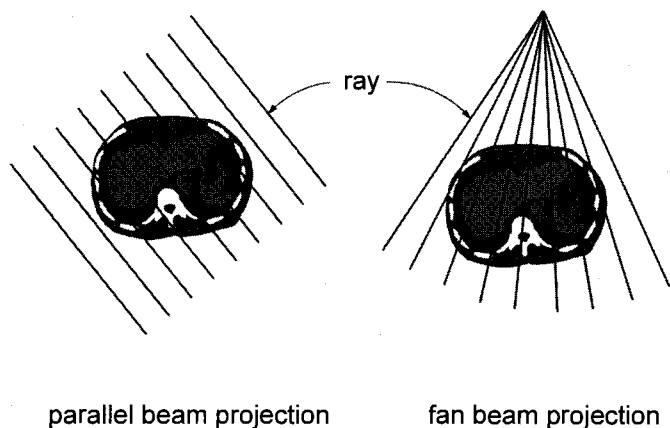


FIGURE 13-3. Computed tomographic (CT) images are produced from a large number of x-ray transmission measurements called rays. A group of rays acquired in a certain geometry is called a *projection* or *view*. Two different geometries have been used in CT, parallel beam projection and fan beam projection, as shown in this figure.

Tomographic Reconstruction

Each ray that is acquired in CT is a transmission measurement through the patient along a line, where the detector measures an x-ray intensity, I_t . The unattenuated intensity of the x-ray beam is also measured during the scan by a reference detector, and this detects an x-ray intensity I_o . The relationship between I_t and I_o is given by the following equation:

$$I_t = I_o e^{-\mu t}$$

where t is the thickness of the patient along the ray and μ is the average linear attenuation coefficient along the ray. Notice that I_t and I_o are machine-dependent values, but the product μt is an important parameter relating to the anatomy of the patient along a given ray. When the equation is rearranged, the measured values I_t and I_o can be used to calculate the parameter of interest:

$$\ln(I_o/I_t) = \mu t$$

where $\ln$ is the natural logarithm (to base e , $e = 2.78 \dots$), t ultimately cancels out, and the value μ for each ray is used in the CT reconstruction algorithm. This computation, which is a preprocessing step performed before image reconstruction, reduces the dependency of the CT image on the machine-dependent parameters, resulting in an image that depends primarily on the patient's anatomic characteristics. This is very much a desirable aspect of imaging in general, and the high clinical utility of CT results, in part, from this feature. By comparison, if a screen-film radiograph is underexposed (I_o is too low) it appears too white, and if it is overexposed (I_o too high) it appears too dark. The density of CT images is independent of I_o , although the noise in the image is affected.

After preprocessing of the raw data, a *CT reconstruction algorithm* is used to produce the CT images. There are numerous reconstruction strategies; however, *filtered backprojection* reconstruction is most widely used in clinical CT scanners. The backprojection method builds up the CT image in the computer by essentially reversing the acquisition steps (Fig. 13-4). During acquisition, attenuation information along a known path of the narrow x-ray beam is integrated by a detector. During backprojection reconstruction, the μ value for each ray is smeared along this same path in the image of the patient. As the data from a large number of rays are

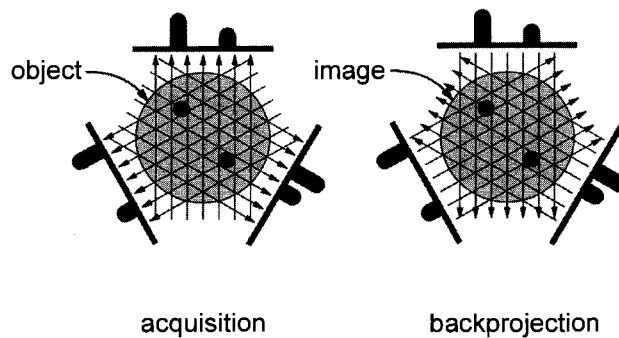


FIGURE 13-4. Data acquisition in computed tomography (CT) involves making transmission measurements through the object at numerous angles around the object (**left**). The process of computing the CT image from the acquisition data essentially reverses the acquisition geometry mathematically (**right**). Each transmission measurement is backprojected onto a digital matrix. After backprojection, areas of high attenuation are positively reinforced through the backprojection process whereas other areas are not, and thus the image is built up from the large collection of rays passing through it.

backprojected onto the image matrix, areas of high attenuation tend to reinforce each other, and areas of low attenuation also reinforce, building up the image in the computer.

13.2 GEOMETRY AND HISTORICAL DEVELOPMENT

Part of understanding an imaging modality involves appreciation of the maturation of that modality. In addition to historical interest, the discussion of the evolution of CT scanners also allows the presentation and reinforcement of several key concepts in CT imaging.

First Generation: Rotate/Translate, Pencil Beam

CT scanners represent a marriage of diverse technologies, including computer hardware, motor control systems, x-ray detectors, sophisticated reconstruction algorithms, and x-ray tube/generator systems. The first generation of CT scanners employed a rotate/translate, pencil beam system (Fig. 13-5). Only two x-ray detectors were used, and they measured the transmission of x-rays through the patient for two different slices. The acquisition of the numerous projections and the multiple rays per projection required that the single detector for each CT slice be physically moved throughout all the necessary positions. This system used parallel ray geometry. Starting at a particular angle, the x-ray tube and detector system translated linearly across the field of view (FOV), acquiring 160 parallel rays across a 24-cm FOV. When the x-ray tube/detector system completed its translation, the whole system was rotated slightly, and then another translation was used to acquire the 160 rays in the next projection. This procedure was repeated until 180 projections were acquired at 1-degree intervals. A total of $180 \times 160 = 28,800$ rays were measured.

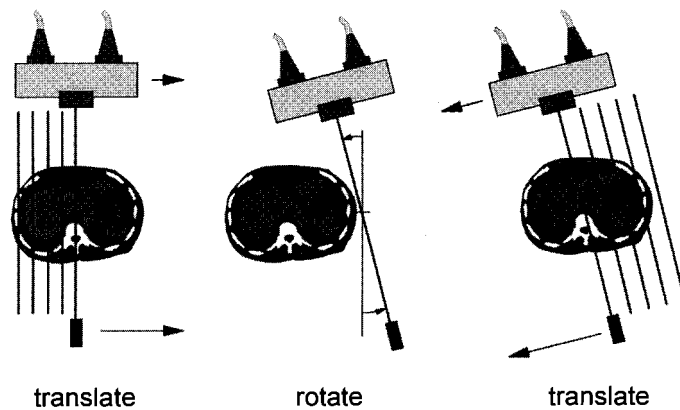


FIGURE 13-5. First-generation (rotate/translate) computed tomography (CT). The x-ray tube and a single detector (per CT slice) translate across the field of view, producing a series of parallel rays. The system then rotates slightly and translates back across the field of view, producing ray measurements at a different angle. This process is repeated at 1-degree intervals over 180 degrees, resulting in the complete CT data set.

As the system translated and measured rays from the thickest part of the head to the area adjacent to the head, a huge change in x-ray flux occurred. The early detector systems could not accommodate this large change in signal, and consequently the patient's head was pressed into a flexible membrane surrounded by a water bath. The water bath acted to bolus the x-rays so that the intensity of the x-ray beam outside the patient's head was similar in intensity to that inside the head. The NaI detector also had a significant amount of "afterglow," meaning that the signal from a measurement taken at one period of time decayed slowly and carried over into the next measurement if the measurements were made temporally too close together.

One advantage of the first-generation CT scanner was that it employed pencil beam geometry—only two detectors measured the transmission of x-rays through the patient. The pencil beam allowed very efficient scatter reduction, because scatter that was deflected away from the pencil ray was not measured by a detector. With regard to scatter rejection, the pencil beam geometry used in first-generation CT scanners was the best.

Second Generation: Rotate/Translate, Narrow Fan Beam

The next incremental improvement to the CT scanner was the incorporation of a linear array of 30 detectors. This increased the utilization of the x-ray beam by 30 times, compared with the single detector used per slice in first-generation systems. A relatively narrow fan angle of 10 degrees was used. In principle, a reduction in scan time of about 30-fold could be expected. However, this reduction time was not realized, because more data ($600 \text{ rays} \times 540 \text{ views} = 324,000 \text{ data points}$) were acquired to improve image quality. The shortest scan time with a second-generation scanner was 18 seconds per slice, 15 times faster than with the first-generation system.

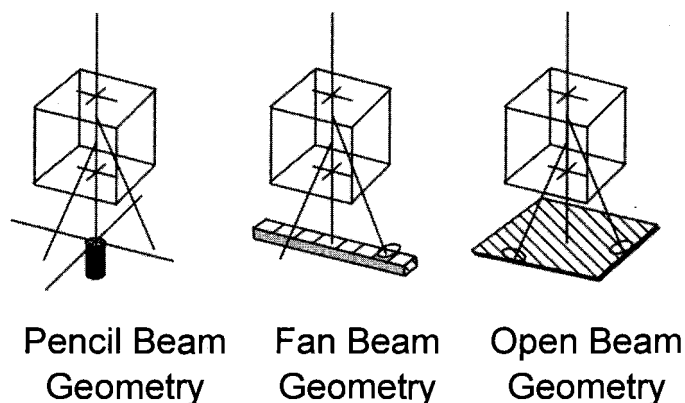


FIGURE 13-6. Pencil beam geometry makes inefficient use of the x-ray source, but it provides excellent x-ray scatter rejection. X-rays that are scattered away from the primary pencil beam do not strike the detector and are not measured. Fan beam geometry makes use of a linear x-ray detector and a divergent fan beam of x-rays. X-rays that are scattered in the same plane as the detector can be detected, but x-rays that are scattered out of plane miss the linear detector array and are not detected. Scattered radiation accounts for approximately 5% of the signal in typical fan beam scanners. Open beam geometry, which is used in projection radiography, results in the highest detection of scatter. Depending on the dimensions and the x-ray energy used, open beam geometries can lead to four detected scatter events for every detected primary photon ($s/p=4$).

Incorporating an array of detectors, instead of just two, required the use of a narrow fan beam of radiation. Although a narrow fan beam provides excellent scatter rejection compared with plain film imaging, it does allow more scattered radiation to be detected than was the case with the pencil beam used in first-generation CT. The difference between pencil beam, fan beam, and open beam geometry in terms of scatter detection is illustrated in Fig. 13-6.

Third Generation: Rotate/Rotate, Wide Fan Beam

The translational motion of first- and second-generation CT scanners was a fundamental impediment to fast scanning. At the end of each translation, the motion of the x-ray tube/detector system had to be stopped, the whole system rotated, and the translational motion restarted. The success of CT as a clinical modality in its infancy gave manufacturers reason to explore more efficient, but more costly, approaches to the scanning geometry.

The number of detectors used in third-generation scanners was increased substantially (to more than 800 detectors), and the angle of the fan beam was increased so that the detector array formed an arc wide enough to allow the x-ray beam to interrogate the entire patient (Fig. 13-7). Because detectors and the associated electronics are expensive, this led to more expensive CT scanners. However, spanning the dimensions of the patient with an entire row of detectors eliminated the need for translational motion. The multiple detectors in the detector array capture the same number of ray measurements in one instant as was required by a complete

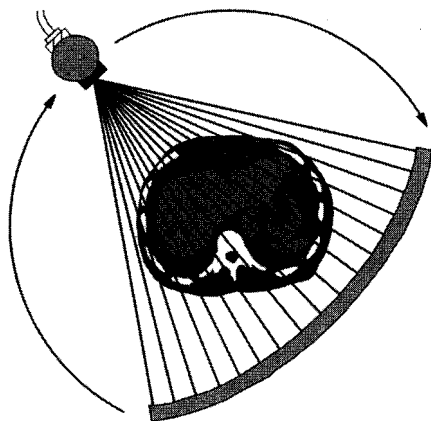


FIGURE 13-7. Third-generation (rotate/rotate) computed tomography. In this geometry, the x-ray tube and detector array are mechanically attached and rotate together inside the gantry. The detector array is long enough so that the fan angle encompasses the entire width of the patient.

translation in the earlier scanner systems. The mechanically joined x-ray tube and detector array rotate together around the patient without translation. The motion of third-generation CT is “rotate/rotate,” referring to the rotation of the x-ray tube and the rotation of the detector array. By elimination of the translational motion, the scan time is reduced substantially. The early third-generation scanners could deliver scan times shorter than 5 seconds. Newer systems have scan times of one-half second.

The evolution from first- to second- and second- to third-generation scanners involved radical improvement with each step. Developments of the fourth- and fifth-generation scanners led not only to some improvements but also to some compromises in clinical CT images, compared with third-generation scanners. Indeed, rotate/rotate scanners are still as viable today as they were when they were introduced in 1975. The features of third- and fourth-generation CT should be compared by the reader, because each offers some benefits but also some trade-offs.

Fourth Generation: Rotate/Stationary

Third-generation scanners suffered from the significant problem of ring artifacts, and in the late 1970s fourth-generation scanners were designed specifically to address these artifacts. It is never possible to have a large number of detectors in perfect balance with each other, and this was especially true 25 years ago. Each detector and its associated electronics has a certain amount of drift, causing the signal levels from each detector to shift over time. The rotate/rotate geometry of third-generation scanners leads to a situation in which each detector is responsible for the data corresponding to a ring in the image (Fig. 13-8). Detectors toward the center of the detector array provide data in the reconstructed image in a ring that is small in diameter, and more peripheral detectors contribute to larger diameter rings.

Third-generation CT uses a fan geometry in which the vertex of the fan is the x-ray focal spot and the rays fan out from the x-ray source to each detector on the detector array. The detectors toward the center of the array make the transmission measurement I_t , while the reference detector that measures I_0 is positioned near the edge of the detector array. If g_1 is the gain of the reference detector, and g_2 is the

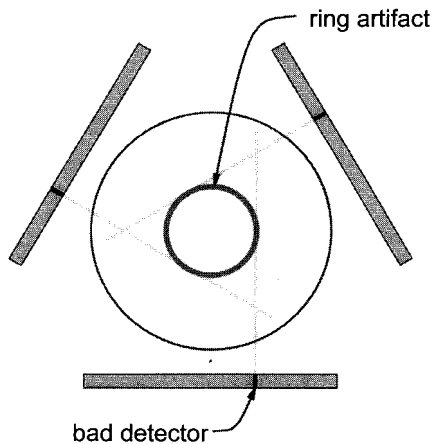


FIGURE 13-8. With third-generation geometry in computed tomography, each individual detector gives rise to an annulus (ring) of image information. When a detector becomes miscalibrated, the tainted data can lead to ring artifacts in the reconstructed image.

gain of the other detector, then the transmission measurement is given by the following equation:

$$\ln(g_1 I_0 / g_2 I_t) = \mu t$$

The equation is true only if the gain terms cancel each other out, and that happens when $g_1 = g_2$. If there is electronic drift in one or both of the detectors, then the gain changes between detectors, so that $g_1 \neq g_2$. So, for third-generation scanners, even a slight imbalance between detectors affects the μt values that are back-projected to produce the CT image, causing the ring artifacts.

Fourth-generation CT scanners were designed to overcome the problem of ring artifacts. With fourth-generation scanners, the detectors are removed from the rotating gantry and are placed in a stationary 360-degree ring around the patient (Fig. 13-9), requiring many more detectors. Modern fourth-generation CT systems use about 4,800 individual detectors. Because the x-ray tube rotates and the detectors are stationary, fourth-generation CT is said to use a rotate/stationary geometry. During acquisition with a fourth-generation scanner, the divergent x-ray beam

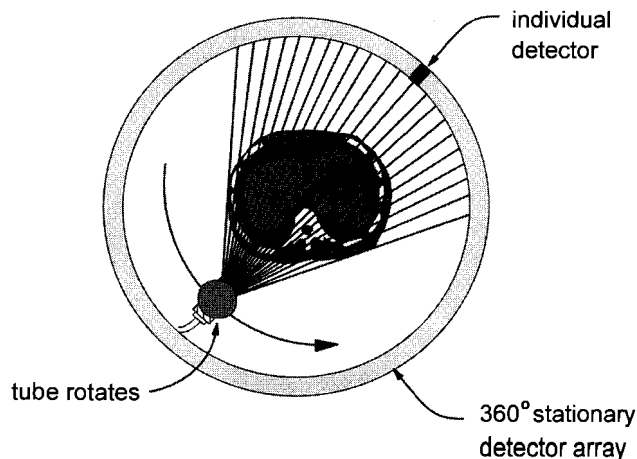


FIGURE 13-9. Fourth-generation (rotate/stationary) computed tomography (CT). The x-ray tube rotates within a complete circular array of detectors, which are stationary. This design requires about six times more individual detectors than a third-generation CT scanner does. At any point during the scan, a divergent fan of x-rays is detected by a group of x-ray detectors.

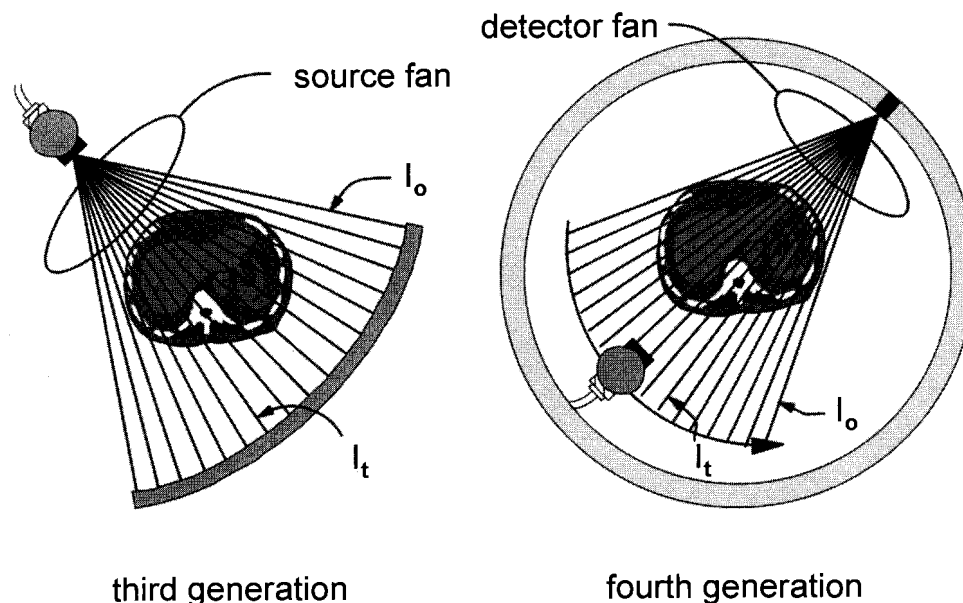


FIGURE 13-10. The fan beam geometry in third-generation computed tomography uses the x-ray tube as the apex of the fan (source fan). Fourth-generation scanners normalize the data acquired during the scan so that the apex of the fan is an individual detector (detector fan). With third-generation scanners, the detectors near the edge of the detector array measure the reference x-ray beam. With fourth-generation scanners, the reference beam is measured by the same detector used for the transmission measurement.

emerging from the x-ray tube forms a fan-shaped x-ray beam. However, the data are processed for fan beam reconstruction with each detector as the vertex of a fan, the rays acquired by each detector being fanned out to different positions of the x-ray source. In the vernacular of CT, third-generation design uses a *source fan*, whereas fourth-generation uses a *detector fan*. The third-generation fan data are acquired by the detector array simultaneously, in one instant of time. The fourth-generation fan beam data are acquired by a single detector over the period of time that is required for the x-ray tube to rotate through the arc angle of the fan. This difference is illustrated in Fig. 13-10. With fourth-generation geometry, each detector acts as its own reference detector. For each detector with its own gain, g , the transmission measurement is calculated as follows:

$$\ln(g I_0 / g I_t) = \mu t$$

Note that the single g term in this equation is guaranteed to cancel out. Therefore, ring artifacts are eliminated in fourth-generation scanners. It should be mentioned, however, that with modern detectors and more sophisticated calibration software, third-generation CT scanners are essentially free of ring artifacts as well.

Fifth Generation: Stationary/Stationary

A novel CT scanner has been developed specifically for cardiac tomographic imaging. This “*cine-CT*” scanner does not use a conventional x-ray tube; instead, a large arc of tungsten encircles the patient and lies directly opposite to the detector ring.

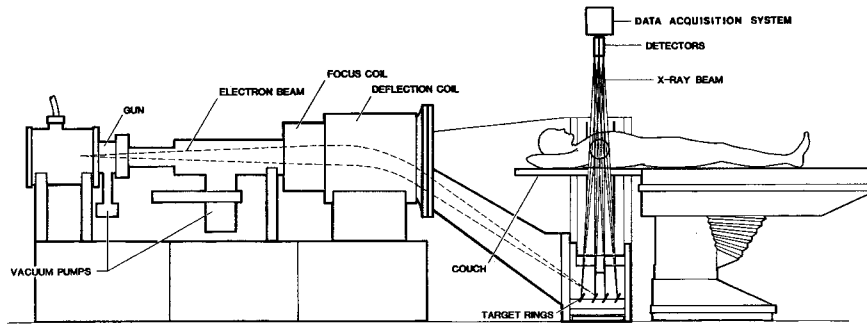


FIGURE 13-11. A schematic diagram of the fifth-generation, Imatron (South San Francisco, CA) cine-CT scanner is shown.

X-rays are produced from the focal track as a high-energy electron beam strikes the tungsten. There are no moving parts to this scanner gantry. The electron beam is produced in a cone-like structure (a vacuum enclosure) behind the gantry and is electronically steered around the patient so that it strikes the annular tungsten target (Fig. 13-11). Cine-CT systems, also called *electron beam scanners*, are marketed primarily to cardiologists. They are capable of 50-msec scan times and can produce fast-frame-rate CT movies of the beating heart.

Sixth Generation: Helical

Third-generation and fourth-generation CT geometries solved the mechanical inertia limitations involved in acquisition of the individual projection data by eliminating the translation motion used in first- and second-generation scanners. However, the gantry had to be stopped after each slice was acquired, because the detectors (in third-generation scanners) and the x-ray tube (in third- and fourth-generation machines) had to be connected by wires to the stationary scanner electronics. The ribbon cable used to connect the third-generation detectors with the electronics had to be carefully rolled out from a cable spool as the gantry rotated, and then as the gantry stopped and began to rotate in the opposite direction the ribbon cable had to be retracted. In the early 1990s, the design of third- and fourth-generation scanners evolved to incorporate *slip ring* technology. A slip ring is a circular contact with sliding brushes that allows the gantry to rotate continually, untethered by wires. The use of slip-ring technology eliminated the inertial limitations at the end of each slice acquisition, and the rotating gantry was free to rotate continuously throughout the entire patient examination. This design made it possible to achieve greater rotational velocities than with systems not using a slip ring, allowing shorter scan times.

Helical CT (also inaccurately called *spiral CT*) scanners acquire data while the table is moving; as a result, the x-ray source moves in a helical pattern around the patient being scanned. Helical CT scanners use either third- or fourth-generation slip-ring designs. By avoiding the time required to translate the patient table, the total scan time required to image the patient can be much shorter (e.g., 30 seconds for the entire abdomen). Consequently, helical scanning allows the use of less contrast agent and increases patient throughput. In some instances the entire scan can be performed within a single breath-hold of the patient, avoiding inconsistent levels of inspiration.

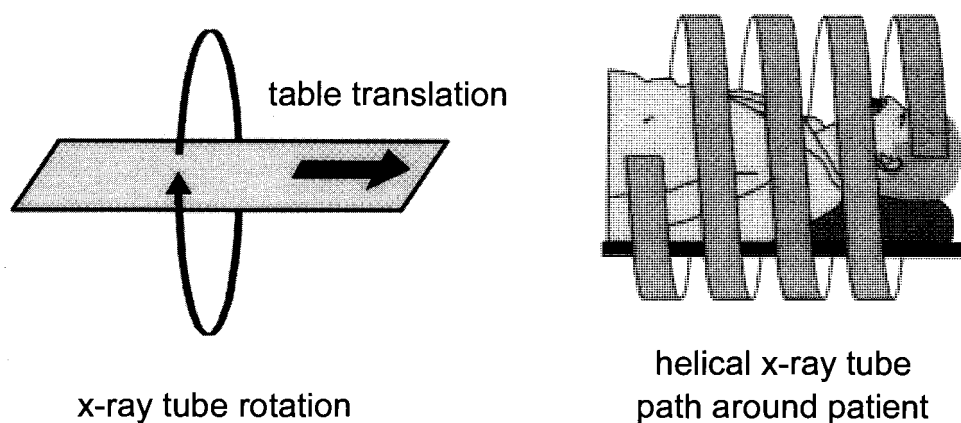


FIGURE 13-12. With helical computed tomographic scanners, the x-ray tube rotates around the patient while the patient and the table are translated through the gantry. The net effect of these two motions results in the x-ray tube traveling in a helical path around the patient.

The advent of helical scanning has introduced many different considerations for data acquisition. In order to produce reconstructions of *planar* sections of the patient, the raw data from the helical data set are interpolated to approximate the acquisition of planar reconstruction data (Fig. 13-12). The speed of the table motion relative to the rotation of the CT gantry is a very important consideration, and the *pitch* is the parameter that describes this relationship (discussed later).

Seventh Generation: Multiple Detector Array

X-ray tubes designed for CT have impressive heat storage and cooling capabilities, although the instantaneous production of x-rays (i.e., x-rays per milliamper-second [mAs]) is constrained by the physics governing x-ray production. An approach to overcoming x-ray tube output limitations is to make better use of the x-rays that are produced by the x-ray tube. When multiple detector arrays are used (Fig. 13-13), the collimator spacing is wider and therefore more of the x-rays that are produced by the x-ray tube are used in producing image data. With conventional, single detector array scanners, opening up the collimator increases the slice thickness, which is good for improving the utilization of the x-ray beam but reduces spatial resolution in the slice thickness dimension. With the introduction of multiple detector arrays, the slice thickness is determined by the detector size and not by the collimator. This represents a major shift in CT technology.

For example, a multiple detector array CT scanner may operate with four contiguous, 5-mm detector arrays and 20-mm collimator spacing. For the same technique (kilovoltage [kV] and mAs), the number of x-rays being detected is four times that of a single detector array with 5-mm collimation. Furthermore, the data set from the 4×5 mm multiple detector array can be used to produce true 5-mm slices, or data from adjacent arrays can be added to produce true 10-, 15-, or even 20-mm slices, all from the same acquisition. The flexibility of CT acquisition protocols and increased efficiency resulting from multiple detector array CT scanners allows better patient imaging; however, the number of parameters involved in the CT acqui-

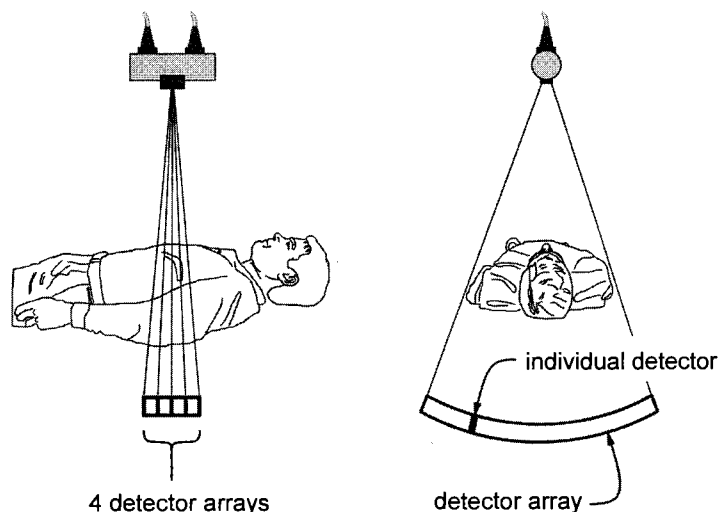


FIGURE 13-13. Multiple detector array computed tomographic (CT) scanners use several, closely spaced, complete detector *arrays*. With no table translation (nonhelical acquisition), each detector array acquires a separate axial CT image. With helical acquisition on a multiple detector array system, table speed and detector pitch can be increased, increasing the coverage for a given period of time.

sition protocol is increased as well. Also with multiple detector arrays, the notion of helical pitch needs to be redefined. This topic is discussed later in the chapter.

13.3 DETECTORS AND DETECTOR ARRAYS

Xenon Detectors

Xenon detectors use high-pressure (about 25 atm) nonradioactive xenon gas, in long thin cells between two metal plates (Fig. 13-14). Although a gaseous detector does not have the same detection efficiency as a solid one, the detector can be made very thick (e.g., 6 cm) to compensate in part for its relatively low density. The metal

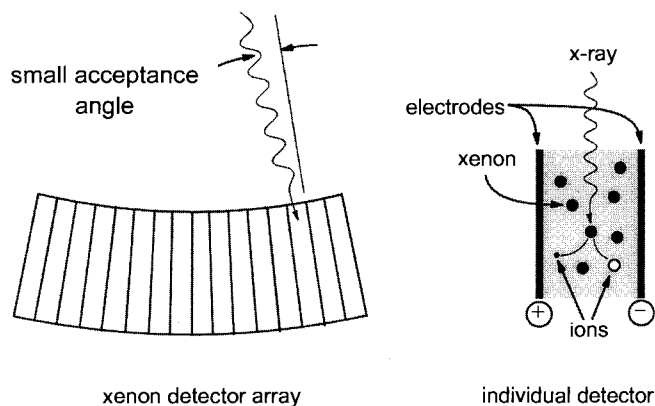


FIGURE 13-14. Xenon detector arrays are a series of highly directional xenon-filled ionization chambers. As x-rays ionize xenon atoms, the charged ions are collected as electric current at the electrodes. The current is proportional to the x-ray fluence.

septa that separate the individual xenon detectors can also be made quite thin, and this improves the geometric efficiency by reducing dead space between detectors. The geometric efficiency is the fraction of primary x-rays exiting the patient that strike active detector elements.

The long, thin ionization plates of a xenon detector are highly directional. For this reason, xenon detectors must be positioned in a fixed orientation with respect to the x-ray source. Therefore, xenon detectors cannot be used for fourth-generation scanners, because those detectors have to record x-rays as the source moves over a very wide angle. Xenon detectors can be used only for third-generation systems.

Xenon detectors for CT are ionization detectors—a gaseous volume is surrounded by two metal electrodes, with a voltage applied across the two electrodes (see Chapter 20). As x-rays interact with the xenon atoms and cause ionization (positive atoms and negative electrons), the electric field (volts per centimeter) between the plates causes the ions to move to the electrodes, where the electronic charge is collected. The electronic signal is amplified and then digitized, and its numerical value is directly proportional to the x-ray intensity striking the detector. Xenon detector technology has been surpassed by solid-state detectors, and its use is now relegated to inexpensive CT scanners.

Solid-State Detectors

A solid-state CT detector is composed of a scintillator coupled tightly to a photodetector. The scintillator emits visible light when it is struck by x-rays, just as in an x-ray intensifying screen. The light emitted by the scintillator reaches the photodetector, typically a photodiode, which is an electronic device that converts light intensity into an electrical signal proportional to the light intensity (Fig. 13-15). This scintillator-photodiode design of solid-state CT detectors is very similar in concept to many digital radiographic x-ray detector systems; however, the performance requirements of CT are slightly different. The detector size in CT is measured in millimeters (typically 1.0×15 mm or 1.0×1.5 mm for multiple detector array scanners), whereas detector elements in digital radiography systems are typi-

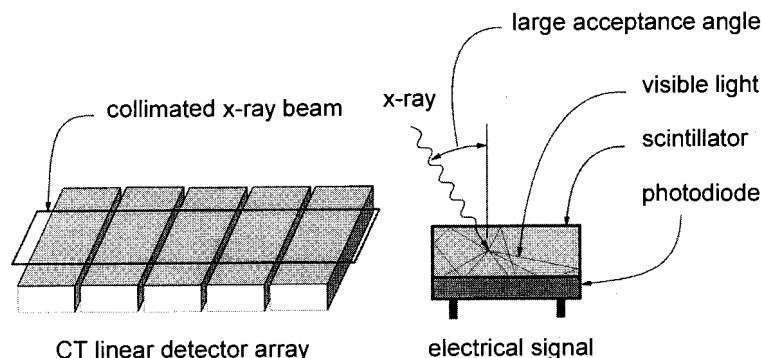


FIGURE 13-15. Solid-state detectors comprise a scintillator coupled to a photodiode. X-rays interact in the scintillator, releasing visible light, which strikes the photodiode and produces an electric signal proportional to the x-ray fluence. For a single CT linear detector array, the detectors are slightly wider than the widest collimated x-ray beam thickness, approximately 12 mm.

cally 0.10 to 0.20 mm on each side. CT requires a very high-fidelity, low-noise signal, typically digitized to 20 or more bits.

The scintillator used in solid-state CT detectors varies among manufacturers, with CdWO_4 , yttrium and gadolinium ceramics, and other materials being used. Because the density and effective atomic number of scintillators are substantially higher than those of pressurized xenon gas, solid-state detectors typically have better x-ray absorption efficiency. However, to reduce crosstalk between adjacent detector elements, a small gap between detector elements is necessary, and this reduces the geometric efficiency somewhat.

The top surface of a solid-state CT detector is essentially flat and therefore is capable of x-ray detection over a wide range of angles, unlike the xenon detector. Solid-state detectors are required for fourth-generation CT scanners, and they are used in most high-tier third-generation scanners as well.

Multiple Detector Arrays

Multiple detector arrays are a set of several linear detector arrays, tightly abutted (Fig. 13-16). The multiple detector array is an assembly of multiple solid-state

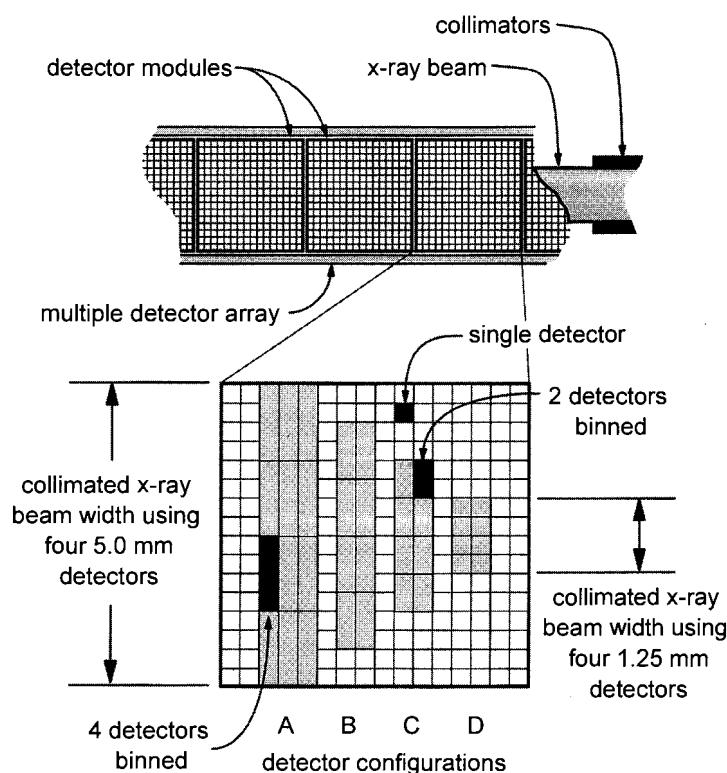


FIGURE 13-16. A multiple detector array composed of multiple detector modules is illustrated in the top panel. In the bottom panel, four different detector configurations are shown, (A–D) each allowing the detection of four simultaneous slices but at different slice widths determined by how many detectors are binned together.

detector array modules. With a traditional single detector array CT system, the detectors are quite wide (e.g., 15 mm) and the adjustable collimator determines slice thickness, typically between 1 and 13 mm. With these systems, the spacing between the collimator blades is adjusted by small motors under computer control. With multiple detector arrays, slice width is determined by the detectors, not by the collimator (although a collimator does limit the beam to the total slice thickness). To allow the slice width to be adjustable, the detector width must be adjustable. It is not feasible, however, to physically change the width of the detector arrays *per se*. Therefore, with multislice systems, the slice width is determined by grouping one or more detector units together. For one manufacturer, the individual detector elements are 1.25 mm wide, and there are 16 contiguous detectors across the module. The detector dimensions are referenced to the scanner's *isocenter*, the point at the center of gantry rotation. The electronics are available for four detector array channels, and one, two, three or four detectors on the detector module can be combined to achieve slices of 4×1.25 mm, 4×2.50 mm, 4×3.75 mm, or 4×5.00 mm. To combine the signal from several detectors, the detectors are essentially wired together using computer-controlled switches. Other manufacturers use the same general approach but with different detector spacings. For example, one manufacturer uses 1-mm detectors everywhere except in the center, where four 0.5-mm-wide detectors are used. Other manufacturers use a gradually increasing spacing, with detector widths of 1.0, 1.5, 2.5, and 5.0 mm going away from the center. Increasing the number of active detector arrays beyond four (used in the example discussed) is a certainty.

Multiple detector array CT scanners make use of solid-state detectors. For a third-generation multiple detector array with 16 detectors in the slice thickness dimension and 750 detectors along each array, 12,000 individual detector elements are used. The fan angle commonly used in third-generation CT scanners is about 60 degrees, so fourth-generation scanners (which have detectors placed around 360 degrees) require roughly six times as many detectors as third-generation systems. Consequently, all currently planned multiple detector array scanners make use of third-generation geometry.

13.4 DETAILS OF ACQUISITION

Slice Thickness: Single Detector Array Scanners

The slice thickness in single detector array CT systems is determined by the physical collimation of the incident x-ray beam with two lead jaws. As the gap between the two lead jaws widens, the slice thickness increases. The width of the detectors in the single detector array places an upper limit on slice thickness. Opening the collimation beyond this point would do nothing to increase slice thickness, but would increase both the dose to the patient and the amount of scattered radiation.

There are important tradeoffs with respect to slice thickness. For scans performed at the same kV and mAs, the number of detected x-ray photons increases *linearly* with slice thickness. For example, going from a 1-mm to a 3-mm slice thickness *triples* the number of detected x-ray photons, and the signal-to-noise ratio (SNR) increases by 73%, since $\sqrt{3} = 1.73$. Increasing the slice thickness from

5 to 10 mm with the same x-ray technique (kV and mAs) *doubles* the number of detected x-ray photons, and the SNR increases by 41%: $\sqrt{2} = 1.41$. Larger slice thicknesses yield better contrast resolution (higher SNR) with the same x-ray techniques, but the spatial resolution in the slice thickness dimension is reduced. Thin slices improve spatial resolution in the thickness dimension and reduce partial volume averaging. For thin slices, the mAs of the study protocol usually is increased to partially compensate for the loss of x-ray photons resulting from the collimation.

It is common to think of a CT image as literally having the geometry of a slab of tissue, but this is not actually the case. The contrast of a small (e.g., 0.5 mm), highly attenuating ball bearing is greater if the bearing is in the center of the CT slice, and the contrast decreases as the bearing moves toward the edges of the slice (Fig. 13-17). This effect describes the *slice sensitivity profile*. For single detector array scanners, the shape of the slice sensitivity profile is a consequence of the finite width of the x-ray focal spot, the penumbra of the collimator, the fact that the image is computed from a number of projection angles encircling the patient, and other minor factors. Furthermore, helical scans have a slightly broader slice sensitivity profile due to translation of the patient during the scan. The *nominal* slice thickness is that which is set on the scanner control panel. Conceptually, the nominal slice is thought of as having a rectangular slice sensitivity profile.

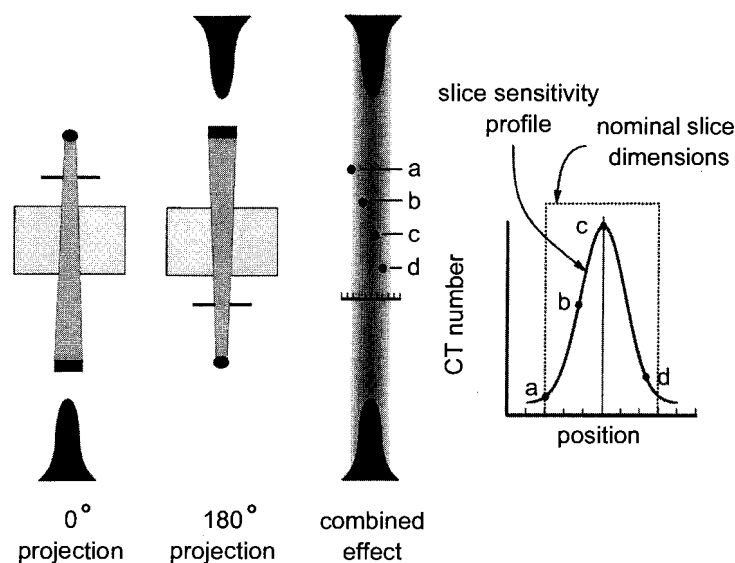


FIGURE 13-17. The slice sensitivity profile is a result of detector width; focal spot size and distribution; collimator penumbra, and the combined influence of all the projection angles. The nominal slice dimensions convey that the computed tomographic (CT) image is of a slab of tissue of finite thickness; however, the same object at different locations in the slice volume (e.g., positions A, B, C, and D) will produce different CT numbers in the image. Consequently, when a small object is placed in the center of the CT slice, it produces greater contrast from background (greater difference in CT number) than when the same object is positioned near the edge of the slice volume.

Slice Thickness: Multiple Detector Array Scanners

The slice thickness of multiple detector array CT scanners is determined not by the collimation, but rather by the width of the detectors in the slice thickness dimension. The width of the detectors is changed by *binning* different numbers of individual detector elements together—that is, the electronic signals generated by adjacent detector elements are electronically summed. Multiple detector arrays can be used both in conventional axial scanning and in helical scanning protocols. In axial scanning (i.e., with no table movement) where, for example, four detector arrays are used, the width of the two center detector arrays almost completely dictates the thickness of the slices. For the two slices at the edges of the scan (detector arrays 1 and 4 of the four active detector arrays), the inner side of the slice is determined by the edge of the detector, but the outer edge is determined either by the collimator penumbra or the outer edge of the detector, depending on collimator adjustment (Fig. 13-18). With a multiple detector array scanner in helical mode, each detector array contributes to every reconstructed image, and therefore the slice sensitivity profile for each detector array needs to be similar to reduce artifacts. To accommodate this condition, it is typical to adjust the collimation so that the focal spot–collimator blade penumbra falls outside the edge detectors. This causes the radiation dose to be a bit higher (especially for small slice widths) in multislice scanners, but it reduces artifacts by equalizing the slice sensitivity profiles between the detector arrays.

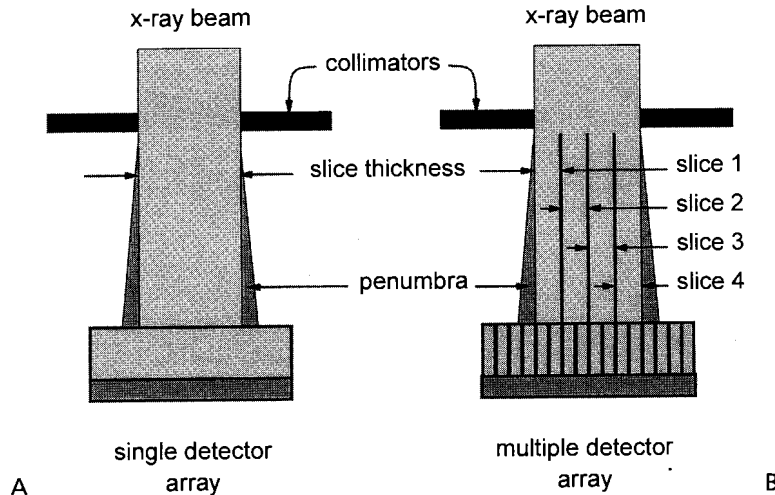


FIGURE 13-18. A: In a single detector array computed tomography (CT) scanner, the width of the CT slice is determined by the collimator. With single detector array scanners, the collimator width is always smaller than the maximum detector width. **B:** Multiple detector array CT scanners acquire several slices simultaneously (four slices shown here), and the x-ray beam collimation determines the outer two edges of the acquired slices. Slice thickness is changed in the single detector array by mechanically changing the collimator position. Slice thickness is determined in the multiple detector array scanner by binning different numbers of detector subunits together and by physically moving the collimator to the outer edges of all four slices.

Detector Pitch and Collimator Pitch

Pitch is a parameter that comes to play when helical scan protocols are used. In a helical CT scanner with one detector array, the pitch is determined by the collimator (collimator pitch) and is defined as:

$$\text{Collimator pitch} = \frac{\text{table movement (mm) per 360-degree rotation of gantry}}{\text{collimator width (mm) at isocenter}}$$

It is customary in CT to measure the collimator and detector widths at the isocenter of the system. The collimator pitch (Fig. 13-19) represents the traditional notion of *pitch*, before the introduction of multiple detector array CT scanners. Pitch is an important component of the scan protocol, and it fundamentally influences radiation dose to the patient, image quality, and scan time. For single detector array scanners, a pitch of 1.0 implies that the number of CT views acquired, when averaged over the long axis of the patient, is comparable to the number acquired with contiguous axial CT. A pitch of less than 1.0 involves overscanning, which may result in some slight improvement in image quality and a higher radiation dose to the patient. CT manufacturers have spent a great deal of developmental effort in optimizing scan protocols for pitches greater than 1.0, and pitches up to 1.5 are commonly used. Pitches with values greater than 1.0 imply some degree of partial scanning along the long axis of the patient. The benefit is faster scan time, less patient motion, and, in some circumstances, use of a smaller volume of contrast agent. Although CT acquisitions around 360 degrees are typical for images of the highest fidelity, the minimum requirement to produce an adequate CT image is a scan of 180 degrees plus the fan angle. With fan angles commonly at about 60 degrees, this means that, at a minimum, $(180 + 60)/360$, or 0.66, of the full circle is required.

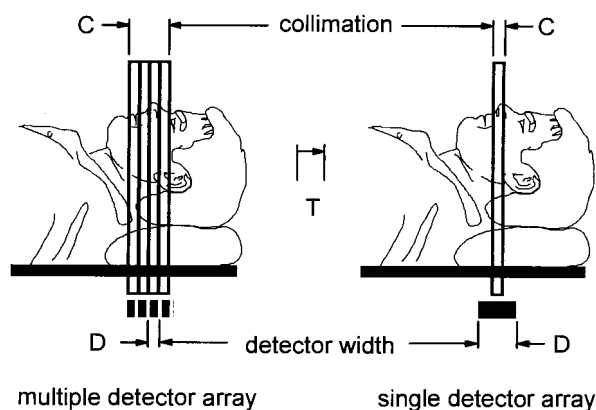


FIGURE 13-19. With a single detector array computed tomographic (CT) scanner, the collimation width is always narrower than the maximum single detector width. A multiple detector array scanner uses collimation that is always wider than a single detector array width. Letting T represent the table translation distance during a 360-degree rotation of the gantry, C would be the collimation width and D would be the detector width. Collimator pitch is defined as T/C , and detector width is defined by T/D . For a multiple detector array CT scanner with four detector arrays, a collimator pitch of 1.0 is equal to a detector pitch of 4.0.

This implies that the upper limit on pitch should be about 1/0.66, or 1.5, because 66% of the data in a 1.5-pitch scan remains contiguous.

Scanners that have multiple detector arrays require a different definition of pitch. The collimator pitch defined previously is still valid, and collimator pitches range between 0.75 and 1.5, as with single detector array scanners. The *detector pitch* (see Fig. 13-19) is also a useful concept for multiple detector array scanners, and it is defined as:

$$\text{Detector pitch} = \frac{\text{table movement (mm) per 360-degree rotation of gantry}}{\text{detector width (mm)}}$$

The need to define detector pitch and collimator pitch arises because beam utilization between single and multiple detector array scanners is different. For a multiple detector array scanner with N detector arrays, the collimator pitch is as follows:

$$\text{Collimator pitch} = \frac{\text{Detector pitch}}{N}$$

For scanners with four detector arrays, detector pitches running from 3 to 6 are used. A detector pitch of 3 for a four-detector array scanner is equivalent to a collimator pitch of 0.75 (3/4), and a detector pitch of 6 corresponds to a collimator pitch of 1.5 (6/4).

13.5 TOMOGRAPHIC RECONSTRUCTION

Rays and Views: The Sinogram

The data acquired for one CT slice can be displayed before reconstruction. This type of display is called a *sinogram* (Fig. 13-20). Sinograms are not used for clinical purposes, but the concepts that they embody are interesting and relevant to understanding tomographic principles. The horizontal axis of the sinogram corresponds to the different rays in each projection. For a third-generation scanner, for example, the horizontal axis of the sinogram corresponds to the data acquired at one instant in time along the length of the detector array. A bad detector in a third-generation scanner would show up as a vertical line on the sinogram. The vertical axis in the sinogram represents each projection angle. A state-of-the-art CT scanner may acquire approximately 1,000 views with 800 rays per view, resulting in a sinogram that is 1,000 pixels tall and 800 pixels wide, corresponding to 800,000 data points. The number of data points (rays/view $\times$ views) acquired by the CT scanner has some impact on the final image quality. For example, first- and second-generation scanners used 28,800 and 324,000 data points, respectively. A modern 512×512 circular CT image contains about 205,000 image pixels, and therefore the ratio between raw data points and image pixels is in the range of about 3.2 to 3.9 (800,000/205,000 = 3.9).

The number of rays used to reconstruct a CT image has a profound influence on the radial component of spatial resolution, and the number of views affects the circumferential component of the resolution (Fig. 13-21). CT images of a simulated object reconstructed with differing numbers of rays (Fig. 13-22) show that reducing the ray sampling results in low-resolution, blurred images. CT images of the simulated object reconstructed with differing numbers of views

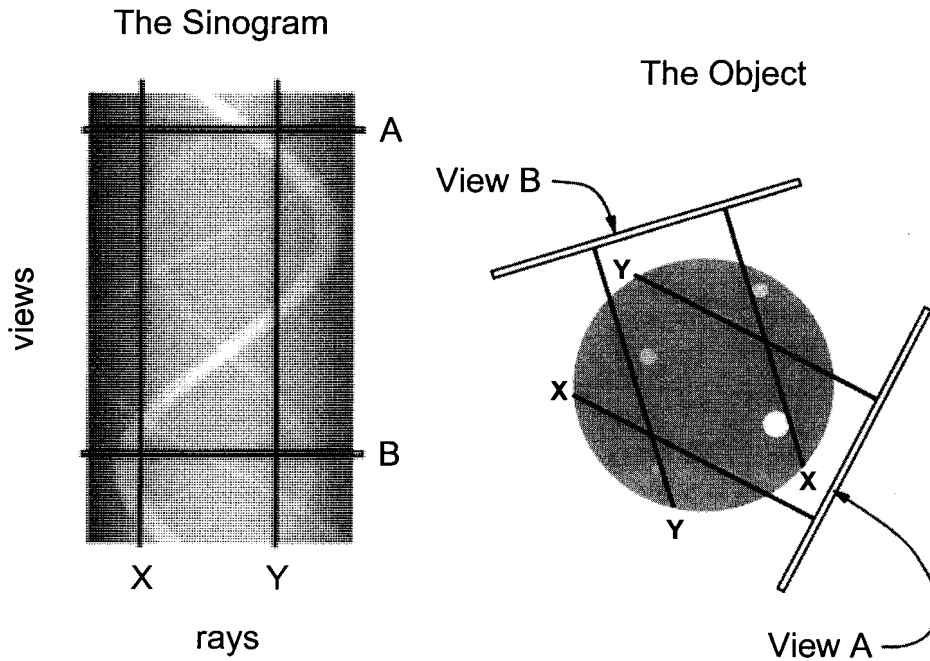


FIGURE 13-20. The sinogram is an image of the raw data acquired by a computed tomographic (CT) scanner before reconstruction. Here, the rays are plotted horizontally and the views are shown on the vertical axis. The sinogram illustrated corresponds to the circular object shown with four circular components. The two views marked A and B on the sinogram correspond to the two views illustrated with the object. During the 360-degree CT acquisition of a particular object, the position of the ray corresponding to that object varies sinusoidally as a function of the view angle. Objects closer to the edge of the field of view produce a sinusoid of high amplitude, and objects closer to the center of rotation have reduced sinusoidal amplitude. An object located exactly at the center of rotation of the gantry produces a shadow that appears vertical on the sinogram.

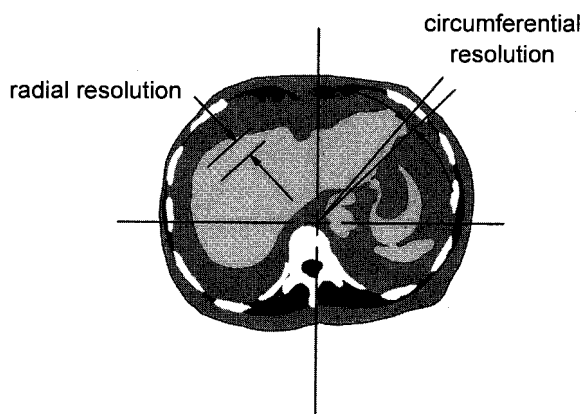


FIGURE 13-21. Unlike screen-film radiography, where the resolution of the image is largely independent of position, the spatial resolution in computed tomography, because of the reconstruction process, has both a radial and circumferential component. The radial component is governed primarily by the spacing and width of the ray data. The circumferential resolution is governed primarily by the number of views used to reconstruct the image.

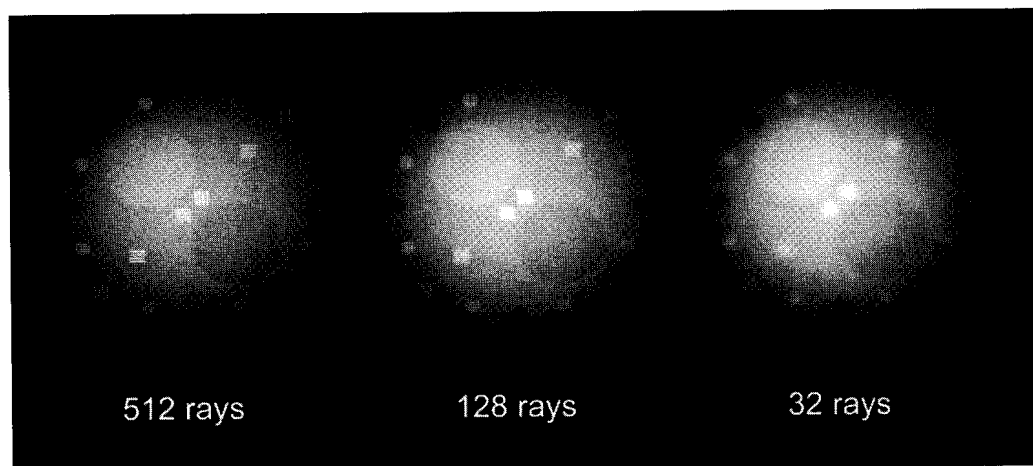


FIGURE 13-22. An image of a test object is shown, reconstructed with 960 views in each image, but with the number of rays differing as indicated. As the number of rays is reduced, the detector aperture was made wider to compensate. With only 32 rays, the detector aperture becomes quite large and the image is substantially blurred as a result.

(Fig. 13-23) show the effect of using too few angular views (*view aliasing*). The sharp edges (high spatial frequencies) of the bar patterns produce radiating artifacts that become more apparent near the periphery of the image. The view aliasing is mild with 240 views and becomes pronounced with 60 views. The presence of high-frequency objects (sharp edges) in the image exacerbates the impact of view aliasing.

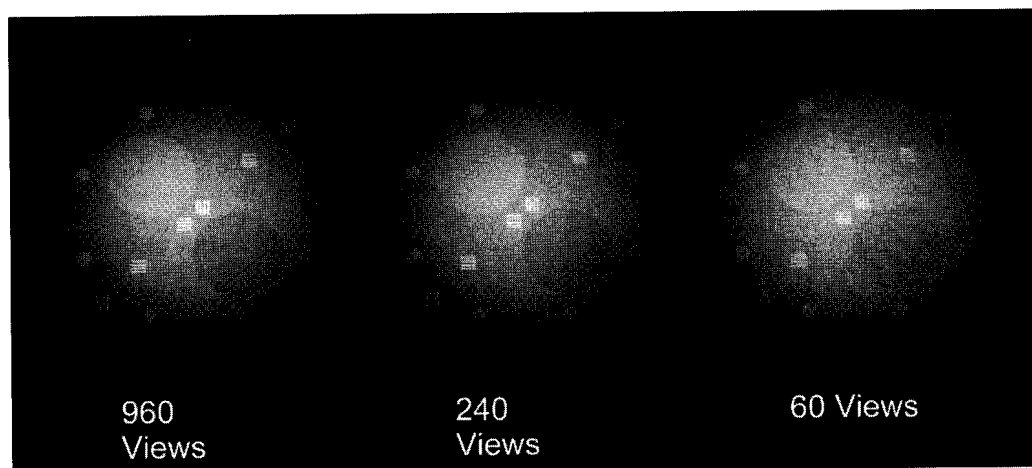


FIGURE 13-23. Each of the three images was reconstructed using 512 rays but differing numbers of views. In going from 960 to 240 views, a slight degree of view aliasing is seen (**center image**). When only 60 views are used (**right image**), the view aliasing is significant. The presence of the aliasing is exacerbated by objects with sharp edges, such as the line pair phantom in the image.

Preprocessing the Data

The raw data acquired by a CT scanner is preprocessed before reconstruction. Numerous preprocessing steps are performed. Electronic detector systems produce a digitized data set that is easily processed by the computer. Calibration data determined from air scans (performed by the technologist or service engineer periodically) provide correction data that are used to adjust the electronic gain of each detector in the array. Variation in geometric efficiencies caused by imperfect detector alignments is also corrected. For example, in fourth-generation scanners (see Fig. 13-9), the x-ray source rotates in an arc around each of the detectors in the array. As the angle and the distance between the detector and the x-ray tube change over this arc, the geometric efficiency of x-ray detection changes. (Similarly, the intensity of sunlight shining on one's palm changes as the wrist is rotated.) These geometric efficiencies are measured during calibration scans and stored in the computer; they are corrected in the preprocessing steps.

After the digital data are calibrated, the logarithm of the signal is computed by the following equation:

$$\ln(I_o/I_t) = \mu t$$

where I_o is the reference data, I_t is the data corresponding to each ray in each view, t is the patient's overall thickness (in centimeters) for that ray, and μ is the linear attenuation coefficient (in cm^{-1}). Note that the data used to calculate the images is linear with respect to the linear attenuation coefficient, μ . The linear attenuation coefficient is determined by the composition and density of the tissue inside each voxel in the patient. Whereas μ is the total attenuation coefficient for each ray (Fig. 13-24), it can be broken down into its components in each small path length Δt :

$$\mu t = \mu_1 \Delta t + \mu_2 \Delta t + \mu_3 \Delta t + \dots + \mu_n \Delta t = \Delta t [\mu_1 + \mu_2 + \mu_3 + \dots + \mu_n]$$

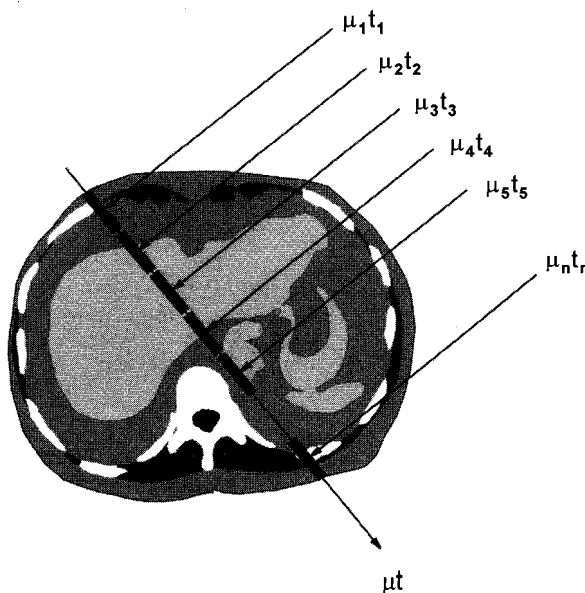


FIGURE 13-24. The measured value of a single ray is equivalent to the attenuation path (t) times the linear attenuation coefficient (μ). The single measurement of μt can be broken up into a series of measurements $\mu_1 t_1$, $\mu_2 t_2$, $\dots$, $\mu_n t_n$, as shown.

The constant Δt factors out, resulting in the following:

$$\mu = \mu_1 + \mu_2 + \mu_3 + \dots + \mu_n$$

Therefore, the reconstructed value in each pixel is the linear attenuation coefficient for the corresponding voxel.

Interpolation (Helical)

Helical CT scanning produces a data set in which the x-ray source has traveled in a helical trajectory around the patient. Present-day CT reconstruction algorithms assume that the x-ray source has negotiated a circular, not a helical, path around the patient. To compensate for these differences in the acquisition geometry, before the actual CT reconstruction the helical data set is interpolated into a series of planar image data sets, as illustrated in Fig. 13-25. Although this interpolation represents an additional step in the computation, it also enables an important feature. With conventional axial scanning, the standard is to acquire contiguous images, which abut one another along the cranial-caudal axis of the patient. With helical scanning, however, CT images can be reconstructed at any position along the length of the scan to within $(\frac{1}{2})$ (pitch) (slice thickness) of each edge of the scanned volume. Helical scanning allows the production of additional overlapping images with *no additional dose* to the patient. Figure 13-26 illustrates the added clinical importance of using interleaved images. The sensitivity of the CT image to objects not centered in the voxel is reduced (as quantified by the slice sensitivity profile), and therefore subtle lesions, which lay between two contiguous images, may be missed. With helical CT scanning, interleaved reconstruction allows the placement of additional images along the patient, so that the clinical examination is almost uniformly sensitive to subtle abnormalities. Interleaved reconstruction adds no additional radiation dose to the patient, but additional time is required to reconstruct the images. Although an increase in the

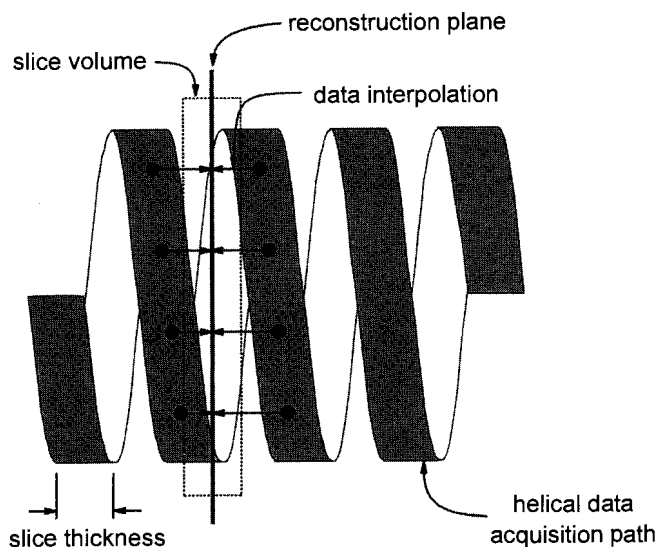


FIGURE 13-25. During helical acquisition, the data are acquired in a helical path around the patient. Before reconstruction, the helical data are interpolated to the reconstruction plane of interest. Interpolation is essentially a weighted average of the data from either side of the reconstruction plane, with slightly different weighting factors used for each projection angle.

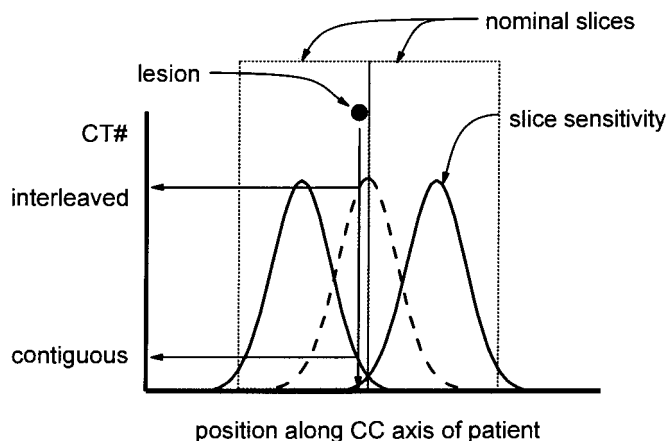


FIGURE 13-26. This figure illustrates the value of interleaved reconstruction. The nominal slice for contiguous computed tomographic (CT) images is illustrated conceptually as two adjacent rectangles; however, the sensitivity of each CT image is actually given by the slice sensitivity profile (solid lines). A lesion that is positioned approximately between the two CT images (black circle) produces low contrast (i.e., a small difference in CT number between the lesion and the background) because it corresponds to low slice sensitivity. With the use of interleaved reconstruction (dashed line), the lesion intersects the slice sensitivity profile at a higher position, producing higher contrast.

image count would increase the interpretation time for traditional side-by-side image presentation, this concern will ameliorate as more CT studies are read by radiologists at computer workstations.

It is important not to confuse the ability to reconstruct CT images at short intervals along the helical data set with the axial resolution itself. The slice thickness (governed by collimation with single detector array scanners and by the detector width in multislice scanners) dictates the actual spatial resolution along the long axis of the patient. For example, images with 5-mm slice thickness can be reconstructed every 1 mm, but this does not mean that 1-mm spatial resolution is achieved. It simply means that the images are sampled at 1-mm intervals. To put the example in technical terms, the *sampling pitch* is 1 mm but the *sampling aperture* is 5 mm. In practice, the use of interleaved reconstruction much beyond a 2:1 interleave yields diminishing returns, except for multiplanar reconstruction (MPR) or 3D rendering applications.

Simple Backprojection Reconstruction

Once the image raw data have been preprocessed, the final step is to use the planar projection data sets (i.e., the preprocessed sinogram) to reconstruct the individual tomographic images. *Iterative reconstruction techniques* are discussed in Chapter 22 and will not be discussed here because they are not used in clinical x-ray CT. As a basic introduction to the reconstruction process, consider Fig. 13-27. Assume that a very simple 2×2 "image" is known only by the projection values. Using algebra ("N equations in M unknowns"), one can solve for the image values in the simple case of a 4-pixel image. A modern CT image contains approx-

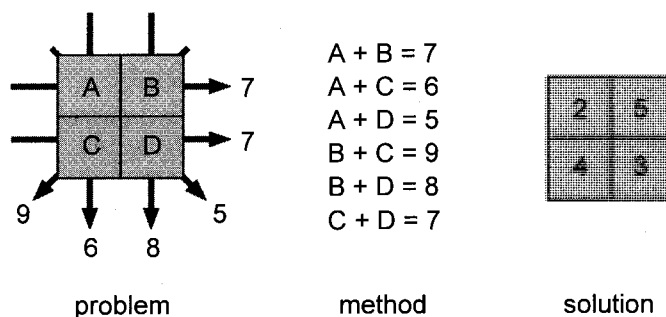


FIGURE 13-27. The mathematical problem posed by computed tomographic (CT) reconstruction is to calculate image data (the pixel values—A, B, C, and D) from the projection values (arrows). For the simple image of four pixels shown here, algebra can be used to solve for the pixel values. With the six equations shown, using substitution of equations, the solution can be determined as illustrated. For the larger images of clinical CT, algebraic solutions become unfeasible, and filtered backprojection methods are used.

imately 205,000 pixels (the circle within a 512×512 matrix) or “unknowns,” and each of the 800,000 projections represent an individual equation. Solving this kind of a problem is beyond simple algebra, and backprojection is the method of choice.

Simple backprojection is a mathematical process, based on trigonometry, which is designed to emulate the acquisition process in reverse. Each ray in each view rep-

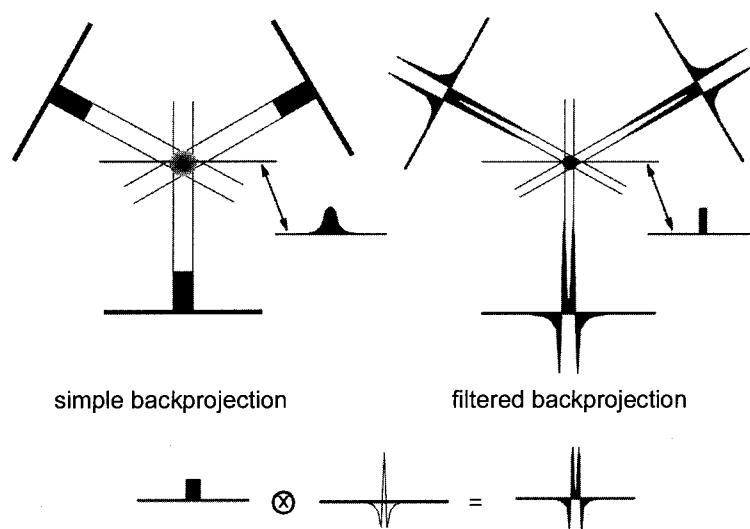


FIGURE 13-28. Simple backprojection is shown on the left; only three views are illustrated, but many views are actually used in computed tomography. A profile through the circular object, derived from simple backprojection, shows a characteristic $1/r$ blurring. With filtered backprojection, the raw projection data are convolved with a convolution kernel and the resulting projection data are used in the backprojection process. When this approach is used, the profile through the circular object demonstrates the crisp edges of the cylinder, which accurately reflects the object being scanned.

resents an individual measurement of μ . In addition to the value of μ for each ray, the reconstruction algorithm also “knows” the acquisition angle and position in the detector array corresponding to each ray. Simple backprojection starts with an empty image matrix (an image with all pixels set to zero), and the μ value from each ray in all views is smeared or backprojected onto the image matrix. In other words, the value of μ is added to each pixel in a line through the image corresponding to the ray’s path.

Simple backprojection comes very close to reconstructing the CT image as desired. However, a characteristic $1/r$ blurring (see later discussion) is a byproduct of simple backprojection. Imagine that a thin wire is imaged by a CT scanner perpendicular to the image plane; this should ideally result in a small point on the image. Rays not running through the wire will contribute little to the image ($\mu = 0$). The backprojected rays, which do run through the wire, will converge at the position of the wire in the image plane, but these projections run from one edge of the reconstruction circle to the other. These projections (i.e., lines) will therefore “radiate” geometrically in all directions away from a point input. If the image gray scale is measured as a function of distance away from the center of the wire, it will gradually diminish with a $1/r$ dependency, where r is the distance away from the point. This phenomenon results in a blurred image of the actual object (Fig. 13-28) when simple backprojection is used. A filtering step is therefore added to correct this blurring, in a process known as *filtered backprojection*.

Filtered Backprojection Reconstruction

In filtered backprojection, the raw view data are mathematically filtered before being backprojected onto the image matrix. The filtering step mathematically reverses the image blurring, restoring the image to an accurate representation of the object that was scanned. The mathematical filtering step involves *convolving* the projection data with a convolution *kernel*. Many convolution kernels exist, and different kernels are used for varying clinical applications such as soft tissue imaging or bone imaging. A typical convolution kernel is shown in Fig. 13-28. The kernel refers to the shape of the filter function in the spatial domain, whereas it is common to perform (and to think of) the filtering step in the frequency domain. Much of the nomenclature concerning filtered backprojection involves an understanding of the frequency domain (which was discussed in Chapter 10). The Fourier transform (FT) is used to convert a function expressed in the spatial domain (millimeters) into the spatial frequency domain (cycles per millimeter, sometimes expressed as mm^{-1}); the inverse Fourier transform (FT^{-1}) is used to convert back. Convolution is an integral calculus operation and is represented by the symbol $\otimes$. Let $p(x)$ represent projection data (in the spatial domain) at a given angle ($p(x)$ is just one horizontal line from a sinogram; see Fig. 13-20), and let $k(x)$ represent the spatial domain kernel as shown in Fig. 13-28. The filtered data in the spatial domain, $p'(x)$, is computed as follows:

$$p'(x) = p(x) \otimes k(x)$$

The difference between filtered backprojection and simple backprojection is the mathematical filtering operation (convolution) shown in this equation above. In filtered backprojection, $p'(x)$ is backprojected onto the image matrix, whereas in

simple backprojection, $p(x)$ is backprojected. The equation can also be performed, quite exactly, in the frequency domain:

$$p'(x) = \text{FT}^{-1}\{\text{FT}[p(x)] \times K(f)\}$$

where $K(f) = \text{FT}[k(x)]$, the kernel in the frequency domain. This equation states that the convolution operation can be performed by Fourier transforming the projection data, *multiplying* (not convolving) this by the frequency domain kernel ($K(f)$), and then applying the inverse Fourier transform on that product to get the filtered data, ready to be backprojected.

Various convolution filters can be used to emphasize different characteristics in the CT image. Several filters, shown in the frequency domain, are illustrated in Fig. 13-29, along with the reconstructed CT images they produced. The Lak filter, named for Dr. Lakshminarayanan, increases the amplitude linearly as a function of frequency and is also called a *ramp filter*. The $1/r$ blurring function in the spatial domain becomes a $1/f$ blurring function in the frequency domain; therefore, multiplying by the Lak filter, where $L(f) = f$, exactly compensates the unwanted $1/f$ blurring, because $1/f \times f = 1$, at all f . This filter works well when there is no noise

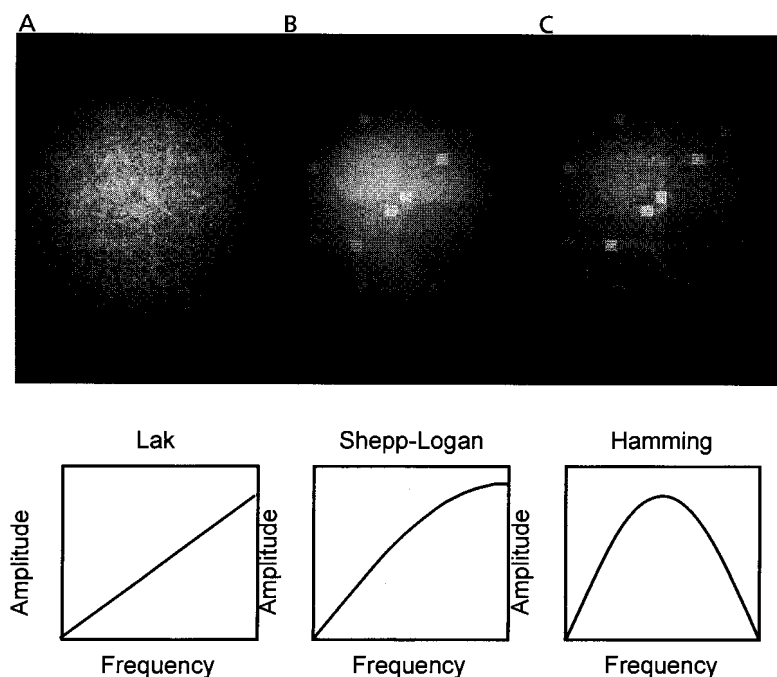


FIGURE 13-29. Three computed tomography (CT) reconstruction filters are illustrated. The plot of the reconstruction filter is shown below, with the image it produced above. **(A)** The Lak is the ideal reconstruction filter in the absence of noise. However, with the quantum noise characteristic in x-ray imaging including CT, the high-frequency components of the noise are amplified and the reconstructed image is consequently very noisy. **(B)** The Shepp-Logan filter is similar to the Lak filter, but roll-off occurs at the higher frequencies. This filter produces an image that is a good compromise between noise and resolution. **(C)** The Hamming filter has extreme high frequency roll-off.

in the data, but in x-ray images there is always x-ray quantum noise, which tends to be more noticeable in the higher frequencies. The Lak filter, as shown in Fig. 13-29A, produces a very noisy CT image. The Shepp-Logan filter in Fig. 13-29B is similar to the Lak filter but incorporates some *roll-off* at higher frequencies, and this reduction in amplification at the higher frequencies has a profound influence in terms of reducing high-frequency noise in the final CT image. The Hamming filter (see Fig. 13-29C) has an even more pronounced high-frequency roll-off, with better high-frequency noise suppression.

Bone Kernels and Soft Tissue Kernels

The reconstruction filters derived by Lakshminarayanan, Shepp and Logan, and Hamming provide the mathematical basis for CT reconstruction filters. In clinical CT scanners, the filters have more straightforward names, and terms such as “bone filter” and “soft tissue filter” are common among CT manufacturers. The term *kernel* is also used. Bone kernels have less high-frequency roll-off and hence accentuate higher frequencies in the image at the expense of increased noise. CT images of bones typically have very high contrast (high signal), so the SNR is inherently quite good. Therefore, these images can afford a slight decrease in SNR ratio in return for sharper detail in the bone regions of the image.

For clinical applications in which high spatial resolution is less important than high contrast resolution—for example, in scanning for metastatic disease in the liver—soft tissue reconstruction filters are used. These kernels have more roll-off at higher frequencies and therefore produce images with reduced noise but lower spatial resolution. The resolution of the images is characterized by the modulation transfer function (MTF); this is discussed in detail in Chapter 10. Figure 13-30 demonstrates two MTFs for state-of-the-art CT scanners. The high-resolution MTF corresponds to use of the bone filter at small FOV, and the standard resolution corresponds to images produced with the soft tissue filter at larger FOV.

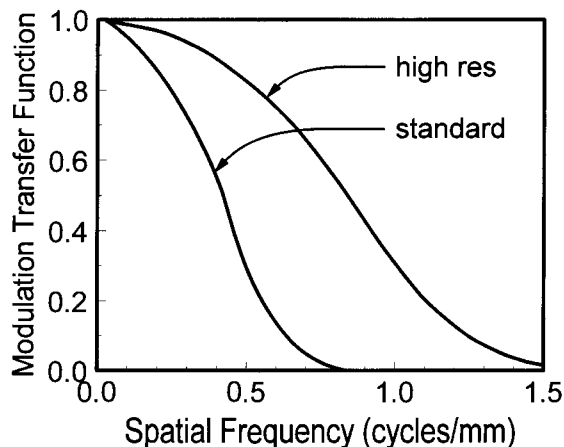


FIGURE 13-30. The modulation transfer function (MTF) for standard and high-resolution clinical computed tomographic scanners are shown. The standard MTF is for a large field of view and a soft tissue reconstruction filter. The high-resolution MTF corresponds to a small field of view and a bone reconstruction filter.

The units for the x-axis in Fig. 13-30 are in cycles per millimeter, and the cut-off frequency is approximately 1.0 cycles/mm. This cutoff frequency is similar to that of fluoroscopy, and it is five to seven times lower than in general projection radiography. CT manufacturers have adapted the unit *cycle/cm*—for example, 1.2 cycles/mm = 12 cycles/cm.

CT Numbers or Hounsfield Units

After CT reconstruction, each pixel in the image is represented by a high-precision floating point number that is useful for computation but less useful for display. Most computer display hardware makes use of integer images. Consequently, after CT reconstruction, but before storing and displaying, CT images are normalized and truncated to integer values. The number $CT(x,y)$ in each pixel, (x,y) , of the image is converted using the following expression:

$$CT(x,y) = 1,000 \frac{\mu(x,y) - \mu_{water}}{\mu_{water}}$$

where $\mu(x,y)$ is the floating point number of the (x,y) pixel before conversion, μ_{water} is the attenuation coefficient of water, and $CT(x,y)$ is the CT number (or Hounsfield unit) that ends up in the final clinical CT image. The value of μ_{water} is about 0.195 for the x-ray beam energies typically used in CT scanning. This normalization results in CT numbers ranging from about -1,000 to +3,000, where -1,000 corresponds to air, soft tissues range from -300 to -100, water is 0, and dense bone and areas filled with contrast agent range up to +3,000.

What do CT numbers correspond to *physically* in the patient? CT images are produced with a highly filtered, high-kV x-ray beam, with an average energy of about 75 keV. At this energy in muscle tissue, about 91% of x-ray interactions are Compton scatter. For fat and bone, Compton scattering interactions are 94% and 74%, respectively. Therefore, CT numbers and hence CT images derive their contrast mainly from the physical properties of tissue that influence Compton scatter. Density (g/cm^3) is a very important discriminating property of tissue (especially in lung tissue, bone, and fat), and the linear attenuation coefficient, μ , tracks linearly with density. Other than physical density, the Compton scatter cross section depends on the electron density (ρ_e) in tissue: $\rho_e = NZ/A$, where N is Avogadro's number (6.023×10^{23} , a constant), Z is the atomic number, and A is the atomic mass of the tissue. The main constituents of soft tissue are hydrogen ($Z = 1$, $A = 1$), carbon ($Z = 6$, $A = 12$), nitrogen ($Z = 7$, $A = 14$), and oxygen ($Z = 8$, $A = 16$). Carbon, nitrogen, and oxygen all have the same Z/A ratio of 0.5, so their electron densities are the same. Because the Z/A ratio for hydrogen is 1.0, the relative abundance of hydrogen in a tissue has some influence on CT number. Hydrogenous tissue such as fat is well visualized on CT. Nevertheless, density (g/cm^3) plays the dominant role in forming contrast in medical CT.

CT numbers are *quantitative*, and this property leads to more accurate diagnosis in some clinical settings. For example, pulmonary nodules that are calcified are typically benign, and the amount of calcification can be determined from the CT image based on the mean CT number of the nodule. Measuring the CT number of a single pulmonary nodule is therefore common practice, and it is an important part of the diagnostic work-up. CT scanners measure bone density with good accuracy, and when phantoms are placed in the scan field along with the patient, quan-

titative CT techniques can be used to estimate bone density, which is useful in assessing fracture risk. CT is also quantitative in terms of linear dimensions, and therefore it can be used to accurately assess tumor volume or lesion diameter.

Computed Tomographic Fluoroscopy Reconstruction

In CT fluoroscopy, the scanner provides pseudo-real time tomographic images that are most commonly used for guidance during biopsies. CT fluoroscopy is similar to conventional CT, with a few added technologic nuances. Like conventional fluoroscopy, CT fluoroscopy provides an image sequence over the same region of tissue, and therefore the radiation dose from repetitive exposure to that tissue is a concern. To address this problem, CT scanners are capable of a CT fluoroscopic mode that uses a low milliamperage setting, typically 20 to 50 mA, whereas 150 to 400 mA is typical for conventional CT imaging.

To provide real-time CT images, special computer hardware is required. Each fluoroscopic CT frame in an imaging sequence is not unique from the preceding CT frame. For example, to achieve 6 CT images per second with a 1-second gantry rotation, the angle scanned during each frame time of $1/6$ second is 60 degrees ($360 \text{ degrees} / 6 = 60 \text{ degrees}$). For CT reconstruction using a full 360-degree projection set, each CT fluoroscopic image uses 17% ($1/6$) new information and 83% ($5/6$) old information (Fig. 13-31). The image acquisition time is still 1 second in this example, but the image is updated 6 times per second. This clever approach provides increased temporal resolution given the physical constraints of backprojection algorithms and the hardware constraints of current technology.

The full data set for each CT fluoroscopic image is not completely re-reconstructed during each frame interval. Using the example of 6 frames per second, after 1 second of operation, consider what occurs in the next $1/6$ second (167 msec): The scanner completes a 60-degree arc, and the projection data acquired during that 167

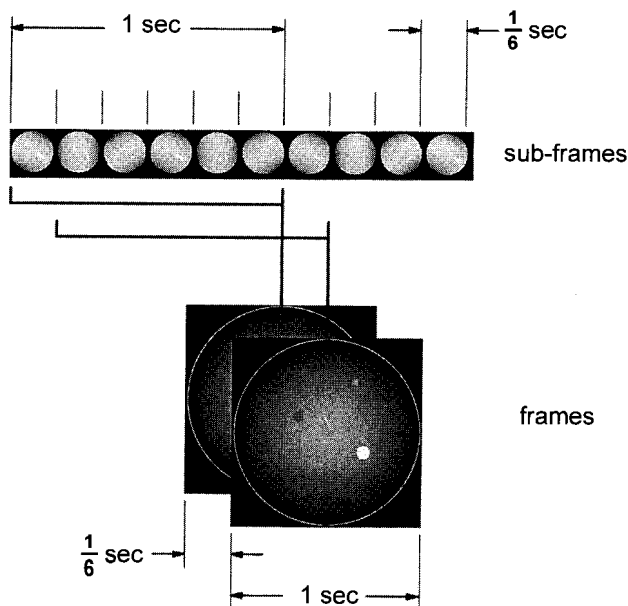


FIGURE 13-31. Data handling in computed tomographic fluoroscopy is illustrated, here using the example of acquisition at 6 frames per second and a 1-second gantry rotation time. During each $1/6$ second, the CT scanner acquires 60 degrees of projection data. Each of these subframes can be filtered and backprojected as shown. A full-frame image is produced by adding six subframes together. Six CT frames are produced each second; however, each CT image corresponds to a scan time of 1 second. For each subsequent CT frame, the newest subframe and the five previous subframes are added.

msec is preprocessed, mathematically filtered, and backprojected to create a CT subframe. This new CT subframe (an image) is added (pixel by pixel) to the five previous subframes, creating a complete CT frame ready for display. Therefore, every 167 msec, a new subframe is created and an old subframe is discarded, so that the most recent six subframes are summed to produce the displayed CT frame. Although 6 frames per second was used in this example, higher frame rates are clinically available.

13.6 DIGITAL IMAGE DISPLAY

Once the CT images from a patient's examination have been reconstructed, the image data must be conveyed to the physician for review and diagnosis. There are some basic postprocessing techniques that are applied to all CT images (i.e., windowing and leveling). For CT studies that have contiguous or interleaving images over a length of the patient, the volume data set lends itself to other forms of postprocessing.

Windowing and Leveling

CT images typically possess 12 bits of gray scale, for a total of 4,096 shades of gray (CT numbers range from $-1,000$ to $+3,095$). Electronic display devices such as computer monitors have the ability to display about 8 bits (256 shades of gray), and laser imagers used for "filming" CT studies have about 8 bits of display fidelity as well. The human eye has incredible dynamic range for adapting to *absolute* light intensities, from moonless nights to desert sunshine. Much of this dynamic range is a result of the eye's ability to modulate the diameter of the light-collecting aperture (the pupil). However, for resolving *relative* differences in gray scale at fixed luminance levels (fixed pupil diameter), as in viewing medical images, the human eye has limited ability (30 to 90 shades of gray), and 6 to 8 bits is considered sufficient for image display.

The 12-bit CT images must be reduced to 8 bits to accommodate most image display hardware. The most common way to perform this postprocessing task (which nondestructively adjusts the image contrast and brightness), is to *window and level* the CT image (Fig. 13-32). The window width (W) determines the contrast of the image, with narrower windows resulting in greater contrast. The level (L) is the CT number at the center of the window. Selection of the values of L and W determines the inflection points P_1 and P_2 , as shown in Fig. 13-32, where $P_1 = L - \frac{1}{2}W$ and $P_2 = L + \frac{1}{2}W$. All the CT numbers lower than P_1 will be saturated to uniform black, and all the CT numbers above P_2 will be saturated to uniform white, and in these saturated areas all image information is lost. It is routine to display CT images using several (two or three) different window and level settings for each image.

Multiplanar Reconstruction

A stack of axial CT images represents anatomic information in three dimensions; however, the axial display is the most common. For studying anatomic features that run along the cranial-caudal dimension of the body, such as the aorta or spinal cord,

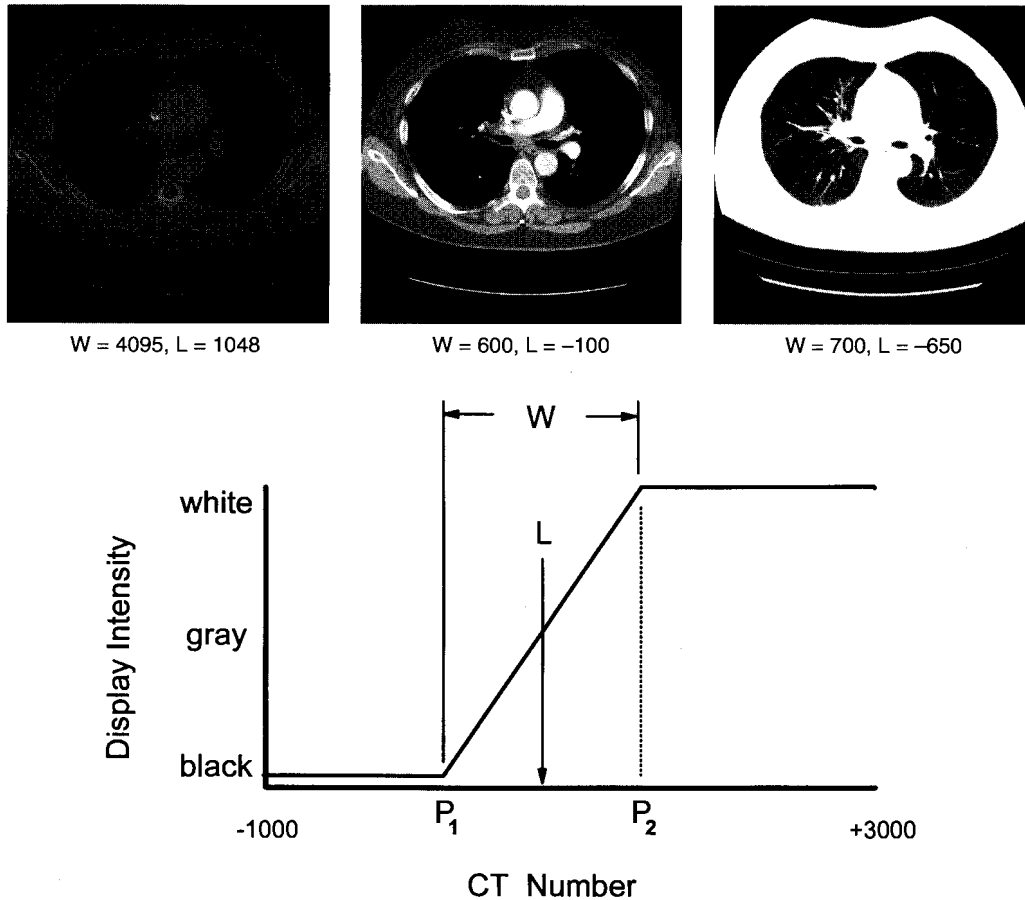


FIGURE 13-32. The concept of how the window and level values are used to manipulate the contrast of the CT image is illustrated. The *level* is the CT number corresponding to the center of the *window*. A narrow window produces a very high contrast image, corresponding to a large slope on the figure. CT numbers below the window (lower than P_1) will be displayed on the image as black; CT numbers above the window (higher than P_2) will be displayed as white. Only CT numbers between P_1 and P_2 will be displayed in a meaningful manner. The thoracic CT images (**top**) illustrate the dramatic effect of changing the window and level settings.

it is sometimes convenient to recompute CT images for display in the sagittal or coronal projections. This is a straightforward technique (Fig. 13-33), and is simply a reformatting of the image data that already exists. Some MPR software packages allow the operator to select a curved reconstruction plane for sagittal or coronal display, to follow the anatomic structure of interest. The spatial resolution of the sagittal or coronal CT images is typically reduced, compared with the axial views. The in-plane pixel dimensions approximate the x-y axis resolution, but the slice thickness limits the z-axis resolution. Sagittal and coronal MPR images combine the x or y dimensions of the CT image with image data along the z-axis, and therefore a mismatch in spatial sampling and resolution occurs. The MPR software compensates for the sampling mismatch by interpolation, so that the MPR images maintain

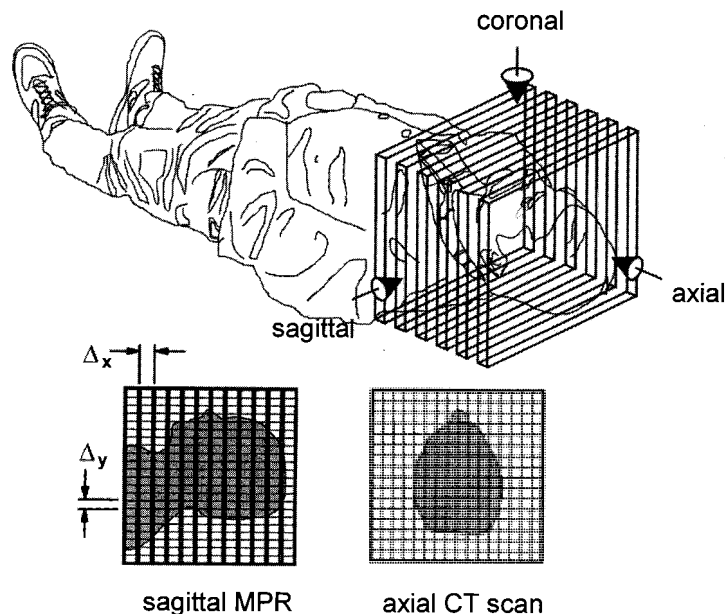


FIGURE 13-33. The coronal, sagittal, and axial projections are illustrated for reference. Both computed tomographic (CT) scans are produced as axial images, and within the image the pixel-to-pixel spacing is small (perhaps 0.5 mm, as a very general number). Coronal or sagittal multiplanar reconstructions (MPRs) can be produced from a series of axial CT scans. For instance, in a sagittal MPR image (**bottom left**), the vertical pixel spacing (Δy) is equivalent to the spacing of the pixels on the axial image. The horizontal spacing (Δx) corresponds to the distance between the reconstructed images. Because Δy is usually much smaller than Δx , it is common to interpolate image data in the horizontal dimension to produce square pixels. For MPR when helical acquisition was used, it is advantageous to reconstruct the images with a significant degree of interleave.

proper aspect ratio (pixel height equals pixel width). However, the system cannot improve the resolution in the z dimension, so MPR images appear a bit blurry along this axis. If it is known before the CT study that MPR images are to be computed, the use of a protocol with thinner CT slices (down to 1 mm) to improve z -axis resolution is warranted.

Three-Dimensional Image Display

Most radiologists are sufficiently facile at viewing series of 2D images to garner a 3D understanding of the patient's anatomy and pathology. In some cases, however, referring physicians prefer to see the CT image data presented in a pseudo-3D manner. Examples would include planning for craniofacial surgery, aortic stent grafting, or radiation treatment. There are two general classes of volume reconstruction techniques: volume rendering techniques and reprojection techniques (Fig. 13-34).

Volume rendering of CT images requires *segmentation*, the identification (by computer alone or with the aid of human intervention) of specific target structures on the 2D image series, before 3D rendering. Segmentation is easiest to perform

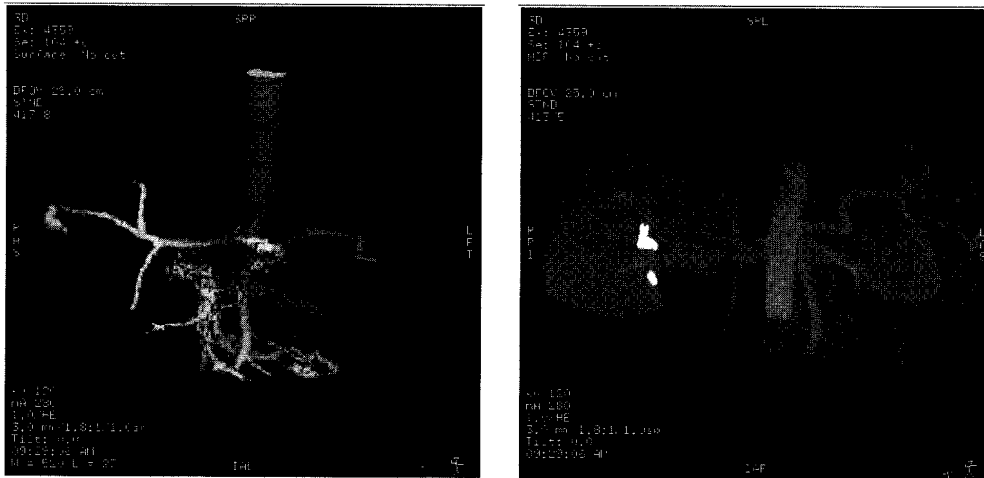


FIGURE 13-34. A volume-rendered image and maximum intensity projection (MIP) image of the aorta and renal arteries are illustrated. Volume-rendered images use surface shading techniques to produce the appearance of a three-dimensional image. The MIP display is similar to a projection radiograph of the volume data set. (Images courtesy of Rick Labonte.)

automatically when there is a large difference in CT number between the target structures and adjacent anatomy (e.g., between bone and soft tissue). The simplest segmentation algorithms allow the user to select a CT number, and the algorithm then assumes that the target structure occupies all pixels with CT numbers greater than the threshold value. Segmentation of the CT data set essentially converts the images to one-bit representations: a “0” if the target structure is not in the pixel and a “1” if it is. Once the target structures (e.g., the facial bones) have been accurately segmented, a computer program then calculates (“renders”) the theoretical appearance of the structure from a specified angle. The software calculates a number of surfaces from the segmented data set, and this process is sometimes called *surface rendering*. Surfaces closest to the viewer obscure surfaces at greater depth. Depending on the sophistication of the software, single or multiple light sources may be factored into the computation to give the appearance of surface shading, color may be added to identify different target structures (i.e., green tumor, white bones, pink spinal cord), and so on. Volume rendering has reached spectacular technologic sophistication in the movie industry, with *Jurassic Park* being the classic example. Medical imaging software has and will continue to benefit from nonmedical rendering advancements, but it will never meet the same degree of success because of the SNR limitations in the data set. The dinosaurs in *Jurassic Park* were calculated from sophisticated mathematical models with perfect SNR, whereas rendering of cranial-facial bones requires a clinical CT data set with imperfect SNR.

Because of SNR limitations in CT imaging, in some studies segmentation, which is required for volume rendering, can be a time-consuming task requiring substantial human intervention, reducing the appeal of this approach. To address this problem, reprojection techniques that do not require segmentation have been developed for the display of volume data. From a specified viewing angle, images are computed from the CT volume data set which are similar geometrically to radiographic projection images. Reprojection techniques simply use ray-tracing soft-

were through the volume data set, from the specified viewing angle. The software can display the normalized sum of the CT numbers from all the voxels through which each ray passed. More commonly, the maximum CT number that each ray encounters is displayed, and this mode is called *maximum intensity projection* (MIP). MIP displays are perhaps not as “three-dimensional” as volume-rendered images, but they can be generated quite reproducibly and completely automatically. To augment the 3D appearance of MIP images, often a series of such images are produced at different viewing angles, and a short movie sequence is generated. For example, a sequence rotating the cervical spine in real time could be displayed. The spatial-temporal cues from the real-time display of these movies can be quite revealing, especially for complicated structures.

Stack Mode Viewing

Many radiology departments are moving to a filmless environment, in which the actual diagnostic interpretation is performed by a radiologist who sits at a computer workstation and views *soft copy* images. One approach is to try to emulate the film alternator to the extent possible, with perhaps four or even six high-resolution monitors clustered together to display the entire CT data set simultaneously. Although this approach seeks to emulate an environment with which most radiologists are familiar, four- or six-monitor workstations are expensive and bulky, and this approach is not designed to take advantage of a key attribute of soft copy display—the computer.

An alternative approach is to simplify the computer hardware, to perhaps two high-resolution monitors, and allow the computer to do the work of the radiologist's neck. In stack mode, a single CT image is displayed and the radiologist selects the image in the study, which is displayed by moving the computer mouse. Often the scout view is presented next to the CT image, with the current image location highlighted. It is possible to simultaneously display two CT images at the same cut plane, such as precontrast and postcontrast images, images at different window and level settings, or images from the previous and the current CT examinations. Certainly many permutations of stack mode display exist, but the essential advantage of this approach is that it is interactive: The radiologist interacts with the computer in real time to visualize the image data as he or she interprets the case, following diagnostic clues from slice to slice. As one clinical example of the benefits of stack mode viewing, a common task in CT is to follow arteries from image to image. On an alternator-like, noninteractive display, the viewer must relocate the vessel spatially, shifting the gaze from image to image. In stack mode display, the viewer's eyes are fixed on the vessel while the various images are displayed under mouse control, allowing the viewer's gaze to follow the vessel as it changes position slightly from image to image.

13.7 RADIATION DOSE

Different x-ray modalities address radiation dose in different ways. For example, in chest radiography it is the entrance *exposure* (not the *dose*) that is the commonly quoted comparison figure. In mammography, the *average glandular dose* is the standard measure of dose. The distribution of radiation dose in CT is markedly differ-

ent than in radiography, because of the unique way in which radiation dose is deposited. There are three aspects of radiation dose in CT that are unique in comparison to x-ray projection imaging. First, because a single CT image is acquired in a highly collimated manner, the volume of tissue that is irradiated by the primary x-ray beam is substantially smaller compared with, for example, the average chest radiograph. Second, the volume of tissue that is irradiated in CT is exposed to the x-ray beam from almost all angles during the rotational acquisition, and this more evenly distributes the radiation dose to the tissues in the beam. In radiography, the tissue irradiated by the entrance beam experiences exponentially more dose than the tissue near the exit surface of the patient. Finally, CT acquisition requires a high SNR to achieve high contrast resolution, and therefore the radiation dose to the slice volume is higher because the techniques used (kV and mAs) are higher. As a rough comparison, a typical PA chest radiograph may be acquired with the use of 120 kV and 5 mAs, whereas a thoracic CT image is typically acquired at 120 kV and 200 mAs.

Dose Measurement

Compton scattering is the principal interaction mechanism in CT, so the radiation dose attributable to scattered radiation is considerable, and it can be higher than the radiation dose from the primary beam. Scattered radiation is not confined to the collimated beam profile as primary x-rays are, and therefore the acquisition of a CT slice delivers a considerable dose from scatter to adjacent tissues, outside the primary beam. Furthermore, most CT protocols call for the acquisition of a series of near-contiguous CT slices over the tissue volume under examination. Take, for example, a protocol in which ten 10-mm CT slices are acquired in the abdomen. The tissue in slice 5 will receive both primary and scattered radiation from its acquisition, but it will also receive the scattered radiation dose from slices 4 and 6, and to a lesser extent from slices 3 and 7, and so on (Fig. 13-35).

The *multiple scan average dose* (MSAD) is the standard for determining radiation dose in CT. The MSAD is the dose to tissue that includes the dose attributable to scattered radiation emanating from all adjacent slices. The MSAD is defined as the average dose, at a particular depth from the surface, resulting from a large series

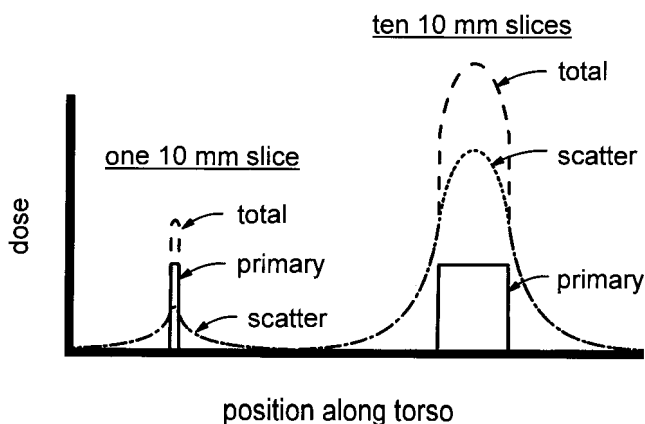


FIGURE 13-35. The radiation dose to the patient is shown as a function of position in the torso along the cranial-caudal dimension. The dose profile for a one 10-mm slice is shown on the left, and the dose profile for ten contiguous 10-mm slices is shown on the right. The contribution (height) of the dose profile from the primary component is the same. However, with a number of adjacent slices, the scattered radiation tails build up and increase the total dose in each CT slice.

of CT slices. The MSAD could be measured directly by placing a small exposure meter at a point in a dose phantom, taking a large series of CT scans of the phantom with the meter in the middle of the slices, and adding the doses from all slices.

An estimate of the MSAD can be accomplished with a single scan by measuring the *CT dose index* (CTDI). It can be shown that the CTDI provides a good approximation to the MSAD when the slices are contiguous. The CTDI measurement protocol seeks to measure the scattered radiation dose from adjacent CT slices in a practical manner. The CTDI is defined by the U.S. Food and Drug Agency (21 CFR § 1 1020.33) as the radiation dose to any point in the patient including the scattered radiation contribution from 7 CT slices in both directions, for a total of 14 slices. In this text, this definition is referred to as the CTDI_{FDA}. One method to measure the CTDI_{FDA} is with many small thermoluminescent dosimeters (TLDs) placed in holes along 14-slice-thickness increments in a Lucite dose phantom. (TLDs are discussed in Chapters 20 and 23.) A single CT image is acquired at the center of the rod containing the TLDs, and the CTDI_{FDA} is determined from this single CT scan by summing the TLD dose measurements.

Medical physicists usually measure the CTDI with the use of a long (100 mm), thin *pencil ionization chamber*. The pencil chamber is long enough to span the width of 14 contiguous 7-mm CT scans and provides a good estimation of the CTDI_{FDA} for 7- and 8-mm slices. A single CT image is acquired at the center of the pencil chamber, and the CTDI is determined from this single CT scan. To calculate the CTDI, all of the energy deposition along the length of the ion chamber is assigned to the thickness of the CT slice:

$$\text{CTDI} = fX/T \times L$$

where X is the measured air kerma (mGy) or exposure (R) to the pencil ion chamber, f is an air-kerma-to-dose (mGy/mGy) or exposure-to-dose (mGy/R or rad/R) conversion factor, L is the length of the pencil ion chamber (i.e., 100 mm), and T is the slice thickness (mm). It can be shown that this method is mathematically equivalent to using a small exposure meter and acquiring all 14 CT scans. The 100-mm pencil ion chamber cannot be used to measure the CTDI_{FDA} for slice thicknesses other than 7 mm. However, it is commonly used for all slice thicknesses, and

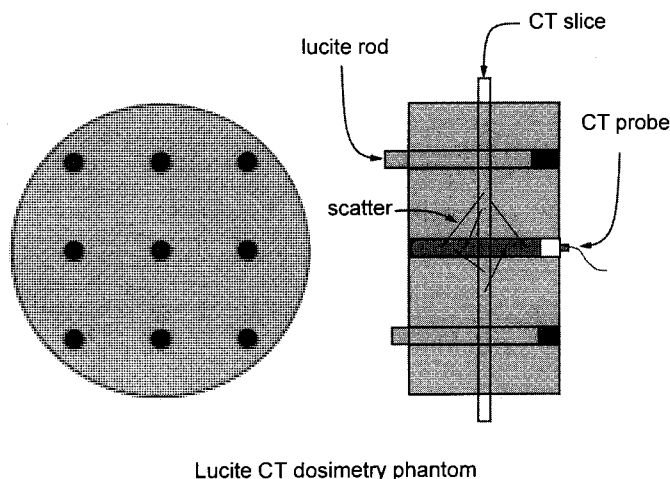


FIGURE 13-36. A diagram of a Lucite phantom, commonly used to determine dose in computed tomography (CT), is shown. The Lucite cylinder has holes drilled out for the placement of the CT pencil chamber. Lucite rods are used to plug in all remaining holes. A CT scan is acquired at the center of the cylinder, and dose measurements are made.

TABLE 13-1. TYPICAL COMPUTED TOMOGRAPHY DOSE INDEX DETERMINED DOSES AT 100 MAS TUBE CURRENT (AVERAGED FROM FOUR MANUFACTURERS DATA)

Peak Kilovoltage (kVp)	Head (16 cm diameter, mGy)	Body (32 cm diameter (mGy))
80	6.4	1.5
100	12.3	3.5
120	17.1	6.4
140	23.4	7.2

these measurements are referred to as the $CTDI_{100}$. The $CTDI_{FDA}$ significantly underestimates the MSAD for small slice thicknesses (e.g., 2 mm), because a significant amount of radiation is scattered beyond seven slice thicknesses. The $CTDI_{100}$ provides much better estimate of the MSAD for thin slices.

There are two standard phantoms that are commercially available for measuring the dose in CT. A 16-cm-diameter Lucite cylinder is used to simulate adult heads and pediatric torsos, and a 32-cm-diameter cylinder is used to simulate the adult abdomen (Fig. 13-36). Holes are drilled parallel to the long axis of the cylinder to allow the placement of the pencil chamber at different positions in the phantom. Lucite rods are used to plug the holes not occupied by the pencil chamber. The CTDI and MSAD are usually measured both at the center of each phantom and at 1 cm from the phantom's surface. The pencil ionization chamber measures air kerma, not dose directly, and the physicist converts this measurement to radiation dose (discussed in Chapter 3) specific to the x-ray beam energy and phantom material used. The CTDI and MSAD are typically expressed as dose to Lucite. For Lucite phantoms in typical CT beams, the conversion factor is approximately 0.893 mGy/mGy (air kerma), or 0.78 rad/roentgen. The dose to soft tissue is about 20% larger than the dose to Lucite. It is important to keep in mind that radiation dose in CT is proportional to the mAs used per slice. At the same kV, doubling of the mAs doubles the dose, and halving the mAs halves the dose. Some typical CTDIs are listed in Table 13-1.

Dose Considerations in Helical Scanning

Helical scanning with a collimator pitch of 1.0 is physically similar to performing a conventional (nonhelical) axial scan with contiguous slices. For CT scanners with multiple detector arrays, the collimator pitch (not the detector pitch) should be used for dose calculations. The dose in helical CT is calculated in exactly the same manner as it is with axial CT, using the CTDI discussed in the previous section; however a correction factor is needed when the pitch is not 1.0:

$$\text{Dose (helical)} = \text{Dose (axial)} \times \frac{1}{\text{Collimator pitch}}$$

For example, with a collimator pitch of 1.5, the dose from helical scanning is 67% that of the dose from conventional scanning, and a collimator pitch of 0.75 corresponds to a helical dose that is 133% of the axial dose (i.e., 33% greater), assuming the same mAs was used (Fig. 13-37). It is important to keep in mind, as mentioned earlier, that the dose changes linearly with the mAs of the study. Because

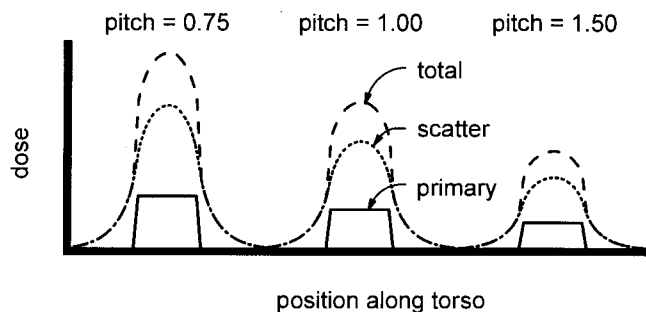


FIGURE 13-37. Dose profiles for three different collimator pitches are illustrated. A pitch of 1.0 corresponds to a dose profile very similar to that of an axial scan. For the same mAs and kV, a (collimator) pitch of 0.75 results in 33% more dose and a pitch of 1.5 corresponds to 66% less dose, compared with a pitch of 1.0. The primary, scatter, and total dose components are shown.

helical scans often use less mAs per 360-degree gantry rotation than axial scans do, the mAs should be factored into the calculation if it differs from the mAs used to measure the CTDI.

Dose in Computed Tomographic Fluoroscopy

Unlike most other CT studies, CT fluoroscopy involves repeated scanning of the same volume of tissue. In order to reduce the dose for these procedures, the mA used for CT fluoroscopic procedures is substantially reduced compared with that used for a regular CT scan. A CT fluoroscopic procedure using 20 mAs per gantry rotation results in a dose that is ten times less than that of a conventional CT scan taken at 200 mAs. The challenge to the radiologist is to perform the procedure, typically a biopsy, using as few seconds of CT fluoroscopy as possible. The dose calculation is straightforward:

$$\text{Dose (fluoro)} = \text{CTDI dose} \times \frac{\text{Time} \times \text{Current}}{\text{mAs of CTDI measurement}}$$

where time is given in seconds and the current (of the CT fluoroscopic examination) in mA. In the example, if 20 seconds of CT fluoroscopy is used at 20 mA and the CTDI (measured at the same kV) was 32 mGy for a 200-mAs study, then the dose would be $32 \text{ mGy} \times (20 \text{ sec} \times 20 \text{ mA}) / 200 \text{ mAs}$, or 64 mGy.

Current Modulation in Computed Tomography

Several CT manufacturers have introduced scanners that are capable of modulating the mA during the scan. In axial cross section, the typical human body is wider than it is thick. The mA modulation technique capitalizes on this fact. The rationale behind this technique is that it takes fewer x-ray photons (lower mA) to penetrate thinner tissue, and more x-ray photons (higher mA) are needed to penetrate thicker projections through the body. The SNR in the final image is related to the number of x-rays that pass through the patient and are detected. However, because of the way in which the filtered backprojection technique combines data from all views,

the high SNR detected in the thinner angular projections is essentially wasted, because the noise levels from the thicker angular projections dominate in the final image. Therefore, a reduction in patient dose can be achieved with little loss in image quality by reducing the mA during acquisition through the thinner tissue projections. The other benefit of this technique is that because the mA is reduced per gantry rotation, x-ray tube loading is reduced and helical scans can be performed for longer periods and with greater physical coverage. The mA modulation technique uses the angular-dependent signal levels (of the detectors) from the first few rotations of the gantry during acquisition (helical and axial) to modulate the mA during subsequent gantry rotations.

13.8 IMAGE QUALITY

Compared with x-ray radiography, CT has significantly worse spatial resolution and significantly better contrast resolution. The MTF, the fundamental measurement of spatial resolution, was shown in Fig. 13-30 for a typical CT scanner; it should be compared with the typical MTF for radiography, discussed in Chapter 10. Whereas the limiting spatial frequency for screen-film radiography is about 7 line pairs (lp) per millimeter and for digital radiography it is 5 lp/mm, the limiting spatial frequency for CT is approximately 1 lp/mm.

It is the contrast resolution of CT that distinguishes the modality: CT has, by far, the best contrast resolution of any clinical x-ray modality. Contrast resolution refers to the ability of an imaging procedure to reliably depict very subtle differences in contrast. It is generally accepted that the contrast resolution of screen-film radiography is approximately 5%, whereas CT demonstrates contrast resolution of about 0.5%. A classic clinical example in which the contrast resolution capability of CT excels is distinguishing subtle soft tissue tumors: The difference in CT number between the tumor and the surrounding tissue may be small (e.g., 20 CT numbers), but because the noise in the CT numbers is smaller (e.g., 3 CT numbers), the tumor is visible on the display to the trained human observer. As is apparent from this example, contrast resolution is fundamentally tied to the SNR. The SNR is also very much related to the number of x-ray quanta used per pixel in the image. If one attempts to reduce the pixel size (and thereby increase spatial resolution) and the dose levels are kept the same, the number of x-rays per pixel is reduced. For example, for the same FOV and dose, changing to a $1,024 \times 1,024$ CT image from a 512×512 image would result in fewer x-ray photons passing through each voxel, and therefore the SNR per pixel would drop. It should be clear from this example that there is a compromise between spatial resolution and contrast resolution. In CT there is a well-established relationship among SNR, pixel dimensions (Δ), slice thickness (T), and radiation dose (D):

$$D \propto \frac{SNR^2}{\Delta^3 T}$$

The clinical utility of any modality lies in its spatial and contrast resolution. In this chapter, many of the factors that affect spatial and contrast resolution of CT have been discussed. Below is a summary of the various factors affecting spatial and contrast resolution in CT.

Factors Affecting Spatial Resolution

Detector pitch: The detector pitch is the center-to-center spacing of the detectors along the array. For third-generation scanners, the detector pitch determines the ray spacing; for fourth-generation scanners, the detector pitch influences the view sampling.

Detector aperture: The detector aperture is the width of the active element of one detector. The use of smaller detectors increases the cutoff (Nyquist) frequency of the image, and it improves spatial resolution at all frequencies.

Number of views: The number of views influences the ability of the CT image to convey the higher spatial frequencies in the image without artifacts (see Fig. 13-23). Use of too few views results in view aliasing, which is most noticeable toward the periphery of the image.

Number of rays: The number of rays used to produce a CT image over the same FOV has a strong influence on spatial resolution (see Fig. 13-22). For a fixed FOV, the number of rays increases as the detector pitch decreases.

Focal spot size: As in any x-ray imaging procedure, larger focal spots cause more geometric unsharpness in the detected image and reduce spatial resolution. The influence of focal spot size is very much related to the magnification of an object to be resolved.

Object magnification: Increased magnification amplifies the blurring of the focal spot. Because of the need to scan completely around the patient in a fixed-diameter gantry, the magnification factors experienced in CT are higher than in radiography. Magnification factors of 2.0 are common, and they can reach 2.7 for the entrant surfaces of large patients.

Slice thickness: The slice thickness is equivalent to the detector aperture in the cranial-caudal axis. Large slice thicknesses clearly reduce spatial resolution in the cranial-caudal axis, but they also reduce sharpness of the edges of structures in the transaxial image. If a linear, high-contrast object such as a contrast-filled vessel runs perfectly perpendicular to the transaxial plane, it will exhibit sharp edges regardless of slice thickness. However, if it traverses through the patient at an angle, its edges will be increasingly blurred with increased slice thickness.

Slice sensitivity profile: The slice sensitivity profile is quite literally the line spread function in the cranial-caudal axis of the patient. The slice sensitivity profile is a more accurate descriptor of the slice thickness.

Helical pitch: The pitch used in helical CT scanning affects spatial resolution, with greater pitches reducing resolution. A larger pitch increases the width of the slice sensitivity profile.

Reconstruction kernel: The shape of the reconstruction kernel has a direct bearing on spatial resolution. Bone filters have the best spatial resolution, and soft tissue filters have lower spatial resolution.

Pixel matrix: The number of pixels used to reconstruct the CT image has a direct influence (for a fixed FOV) on spatial resolution; however no CT manufacturer compromises on this parameter in a way that reduces resolution. In some instances the pixel matrix may be reduced. For example, some vendors reconstruct to a 256×256 pixel matrix for CT fluoroscopy to achieve real-time performance. Also, in soft copy viewing, if the number of CT images displayed on a monitor exceeds the pixel resolution of the display hardware, the software downscans the CT images, reducing the spatial resolution. For example, a $1,024 \times 1,024$ pixel workstation can display only four 512×512 CT images at full resolution.

Patient motion: If there is involuntary motion (e.g., heart) or motion resulting from patient noncompliance (e.g., respiratory), the CT image will experience blurring proportional to the distance of the motion during the scan.

Field of view: The FOV influences the physical dimensions of each pixel. A 10-cm FOV in a 512×512 matrix results in pixel dimensions of approximately 0.2 mm, and a 35-cm FOV produces pixel widths of about 0.7 mm.

Factors Affecting Contrast Resolution

mAs: The mAs directly influence the number of x-ray photons used to produce the CT image, thereby affecting the SNR and the contrast resolution. Doubling of the mAs of the study increases the SNR by $\sqrt{2}$ or 41%, and the contrast resolution consequently improves.

Dose: The dose increases linearly with mAs per scan, and the comments for mAs apply here.

Pixel size (FOV): If patient size and all other scan parameters are fixed, as FOV increases, pixel dimensions increase, and the number of x-rays passing through each pixel increases.

Slice thickness: The slice thickness has a strong (linear) influence on the number of photons used to produce the image. Thicker slices use more photons and have better SNR. For example, doubling of the slice thickness doubles the number of photons used (at the same kV and mAs), and increases the SNR by $\sqrt{2}$, or 41%.

Reconstruction filter: Bone filters produce lower contrast resolution, and soft tissue filters improve contrast resolution.

Patient size: For the same x-ray technique, larger patients attenuate more x-rays, resulting in detection of fewer x-rays. This reduces the SNR and therefore the contrast resolution.

Gantry rotation speed: Most CT systems have an upper limit on mA, and for a fixed pitch and a fixed mA, faster gantry rotations (e.g., 1 second compared with 0.5 second) result in reduced mAs used to produce each CT image, reducing contrast resolution.

13.9 ARTIFACTS

Beam Hardening

Like all medical x-ray beams, CT uses a polyenergetic x-ray spectrum, with energies ranging from about 25 to 120 keV. Furthermore, x-ray attenuation coefficients are energy dependent: After passing through a given thickness of patient, lower-energy x-rays are attenuated to a greater extent than higher-energy x-rays are. Therefore, as the x-ray beam propagates through a thickness of tissue and bone, the shape of the spectrum becomes skewed toward the higher energies (Fig. 13-38). Consequently, the average energy of the x-ray beam becomes greater ("harder") as it passes through tissue. Because the attenuation of bone is greater than that of soft tissue, bone causes more beam hardening than an equivalent thickness of soft tissue (Fig. 13-39).

The beam-hardening phenomenon induces artifacts in CT because rays from some projection angles are hardened to a differing extent than rays from other angles, and this confuses the reconstruction algorithm. A calculated example is shown in Fig. 13-40A. The most common clinical example of beam hardening

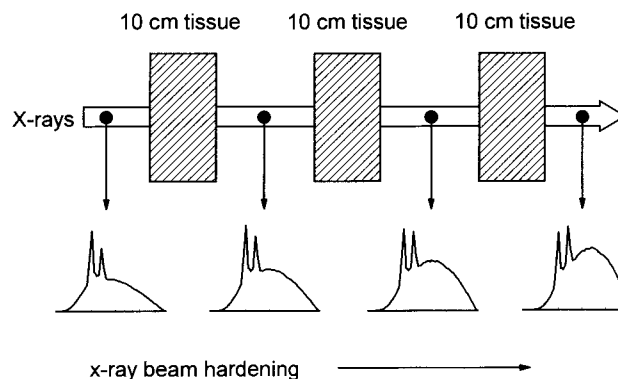


FIGURE 13-38. The nature of x-ray beam hardening is illustrated. As a spectrum of x-rays (**lower graphs**) passes layers of tissue, the lower-energy photons in the x-ray spectrum are attenuated to a greater degree than the higher-energy components of the spectrum. Therefore, as the spectrum passes through increasing thickness of tissue, it becomes progressively skewed toward the higher-energy x-rays in that spectrum. In the vernacular of x-ray physics, a higher-energy spectrum is called a “harder” spectrum; hence the term *beam hardening*.

occurs between the petrous bones in the head, where a spider web–like artifact connects the two bones on the image.

Most CT scanners include a simple beam-hardening correction algorithm that corrects for beam hardening based on the relative attenuation of each ray, and this helps to some extent. More sophisticated two-pass beam-hardening correction algorithms have been implemented by some manufacturers, and these correction techniques can substantially reduce artifacts resulting from beam hardening. In a two-pass algorithm, the image is reconstructed normally in the first pass, the path length

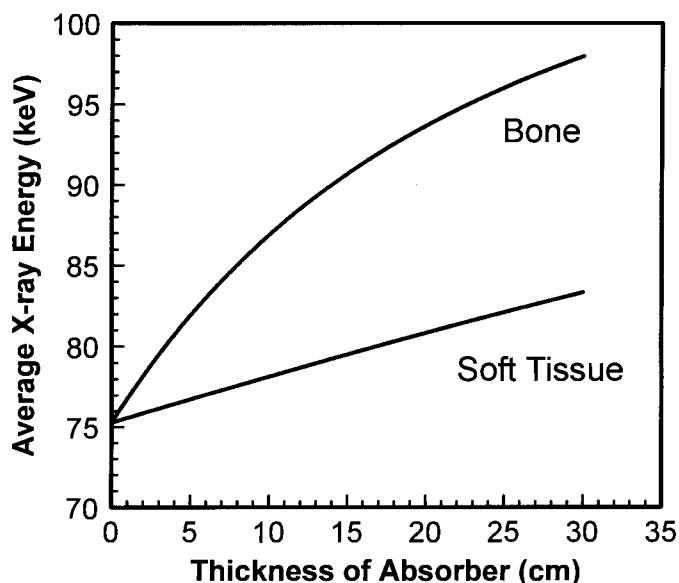


FIGURE 13-39. Using a 120-kV x-ray spectrum filtered by 7 mm of aluminum as the incident x-ray spectrum, the average x-ray energy of the spectrum is shown as a function of the thickness of the absorber through which the spectrum propagates. Beam hardening caused by soft tissue and by bone is illustrated. Bone causes greater beam hardening because of its higher attenuation coefficient (due to increased density and higher average atomic number).

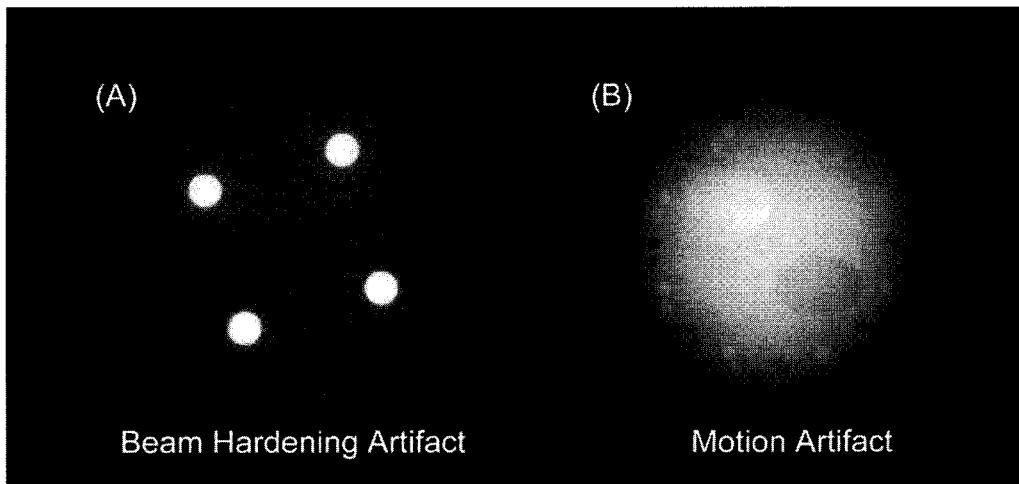


FIGURE 13-40. A: A beam-hardening artifact caused by four calcium-filled cylinders is illustrated. **B:** Artifacts due to motion are shown.

that each ray transits through bone and soft tissue is determined from the first-pass image, and then each ray is compensated for beam hardening based on the known bone and soft tissue thicknesses it has traversed. A second-pass image reconstruction is performed using the corrected ray values.

Motion Artifacts

Motion artifacts occur when the patient moves during the acquisition. Small motions cause image blurring, and larger physical displacements during CT image acquisition produce artifacts that appear as double images or image ghosting. If motion is suspected in a CT image, the adjacent CT scans in the study may be evaluated to distinguish fact from artifact. In some cases, the patient needs to be re-scanned. An example of a motion artifact is shown in Fig. 13-40B.

Partial Volume Averaging

The CT number in each pixel is proportional to the average μ in the corresponding voxel. For voxels containing all one tissue type (e.g., all bone, all liver), μ is representative of that tissue. Some voxels in the image, however, contain a mixture of different tissue types. When this occurs, for example with bone and soft tissue, the μ is not representative of either tissue but instead is a weighted average of the two different μ values. Partial volume averaging is most pronounced for softly rounded structures that are almost parallel to the CT slice. The most evident example is near the top of the head, where the cranium shares a substantial number of voxels with brain tissue, causing details of the brain parenchyma to be lost because the large μ of bone dominates. This situation is easily recognizable and therefore seldom leads to misdiagnosis. Partial volume artifacts can lead to misdiagnosis when the presence of adjacent anatomic structures is not suspected (Fig. 13-41).

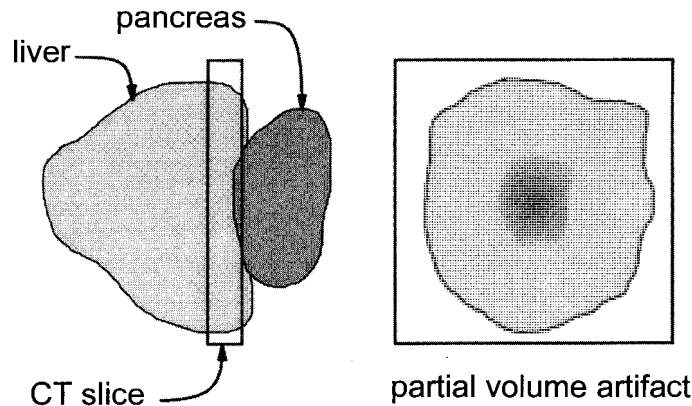


FIGURE 13-41. A partial volume artifact occurs when the computed tomographic slice interrogates a slab of tissue containing two or more different tissue types. Many partial volume artifacts are obvious (e.g., with bone), but occasionally a partial volume artifact can mimic pathologic conditions.

There are several approaches to reducing partial volume artifacts, and the obvious *a priori* approach is to use thinner CT slices. When a suspected partial volume artifact occurs with a helical study and the raw scan data is still available, it is sometimes necessary to use the raw data to reconstruct additional CT images at different positions. Many institutions make use of interleaved reconstructions (e.g., 5-mm slices every 2.5 mm) as a matter of routine protocol.

SUGGESTED READING

- Cho ZH, Jones JP, Singh M. *Foundations of medical imaging*. New York: John Wiley and Sons, 1993.
- Goldman LW, Fowlkes JB. *Syllabus from the 2000 categorical course in diagnostic radiology physics: CT and US cross sectional imaging*. Oak Brook, IL: RSNA Publications, 2000.
- Herman GT. *Image reconstruction from projections: the fundamentals of computerized tomography*. New York: Academic Press, 1980.
- Huda W, Atherton JV. Energy imparted in computed tomography. *Med Phys* 1994;22:1263–1269.
- Sprawls P. AAPM tutorial: CT image detail and noise. *Radiographics* 1992;12:1041–1046.

NUCLEAR MAGNETIC RESONANCE

Nuclear magnetic resonance (NMR) is the spectroscopic study of the magnetic properties of the *nucleus* of the atom. The protons and neutrons of the nucleus have a *magnetic* field associated with their nuclear spin and charge distribution. *Resonance* is an energy coupling that causes the individual nuclei, when placed in a strong external magnetic field, to selectively absorb, and later release, energy unique to those nuclei and their surrounding environment. The detection and analysis of the NMR signal has been extensively studied since the 1940s as an analytic tool in chemistry and biochemistry research. NMR is not an imaging technique but rather a method to provide spectroscopic data concerning a sample placed in the device. In the early 1970s, it was realized that magnetic field gradients could be used to localize the NMR signal and to generate images that display magnetic properties of the proton, reflecting clinically relevant information. As clinical imaging applications increased in the mid-1980s, the “nuclear” connotation was dropped, and magnetic resonance imaging (MRI), with a plethora of associated acronyms, became commonly accepted in the medical community.

MRI is a rapidly changing and growing image modality. The high contrast sensitivity to soft tissue differences and the inherent safety to the patient resulting from the use of nonionizing radiation have been key reasons why MRI has supplanted many CT and projection radiography methods. With continuous improvements in image quality, acquisition methods, and equipment design, MRI is the modality of choice to examine anatomic and physiologic properties of the patient. There are drawbacks, including high equipment and siting costs, scan acquisition complexity, relatively long imaging times, significant image artifacts, and patient claustrophobia problems.

The expanding use of MRI demands an understanding of the underlying MR principles to glean the most out of the modality. This chapter reviews the basic properties of magnetism, concepts of resonance, and generation of tissue contrast through specific pulse sequences. Details of image formation, characteristics, and artifacts, as well as equipment operation and biologic effects, are discussed in Chapter 15.

14.1 MAGNETIZATION PROPERTIES

Magnetism

Magnetism is a fundamental property of matter; it is generated by moving charges, usually electrons. Magnetic properties of materials result from the organization and

motion of the electrons in either a random or a nonrandom alignment of magnetic “domains,” which are the smallest entities of magnetism. Atoms and molecules have electron orbitals that can be paired (an even number of electrons cancels the magnetic field) or unpaired (the magnetic field is present). Most materials do not exhibit overt magnetic properties, but one notable exception is the permanent magnet, in which the individual domains are aligned in one direction.

Magnetic susceptibility describes the extent to which a material becomes magnetized when placed in a magnetic field. The induced internal magnetization can oppose the external magnetic field and lower the local magnetic field surrounding the material. On the other hand, the internal magnetization can form in the same direction as the applied magnetic field and increase the local magnetic field. Three categories of susceptibility are defined: *diamagnetic*, *paramagnetic*, and *ferromagnetic*. Diamagnetic materials have slightly negative susceptibility and oppose the applied magnetic field. Examples of diamagnetic materials are calcium, water, and most organic materials (chiefly owing to the diamagnetic characteristics of carbon and hydrogen). Paramagnetic materials have slightly positive susceptibility and enhance the local magnetic field, but they have no measurable self-magnetism. Examples of paramagnetic materials are molecular oxygen (O_2), some blood degradation products, and gadolinium-based contrast agents. Ferromagnetic materials are “superparamagnetic”—that is, they augment the external magnetic field substantially. These materials can exhibit “self-magnetism” in many cases. Examples are iron, cobalt, and nickel.

Unlike the monopole electric charges from which they are derived, magnetic fields exist as dipoles, where the north pole is the origin of the magnetic field lines and the south pole is the return (Fig. 14-1A). One pole cannot exist without the other. As with electric charges, “like” magnetic poles repel and “opposite” poles attract. The *magnetic field strength*, B , (also called the magnetic flux density) can be

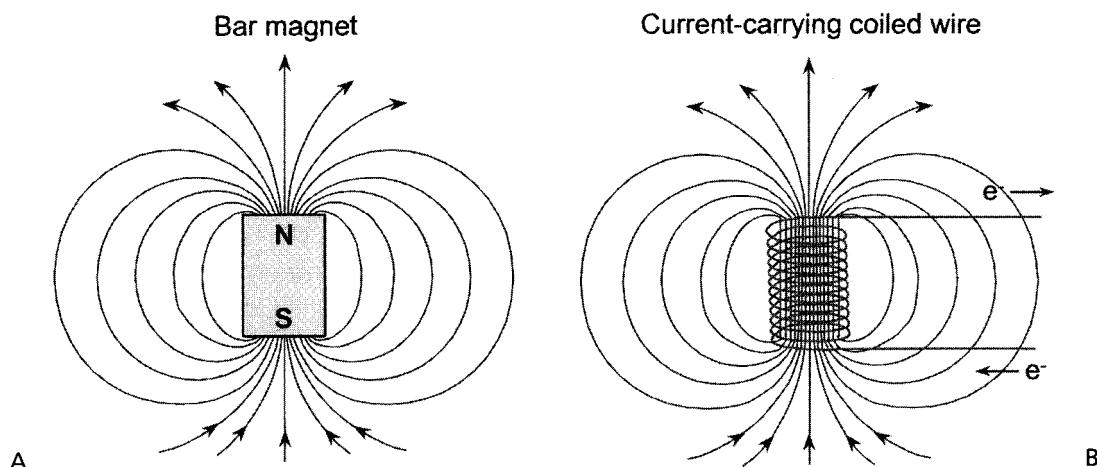


FIGURE 14-1. A: The magnetic field has two poles, with magnetic field lines emerging from the north pole (N) and returning to the south pole (S), as illustrated by a simple bar magnet. **B:** A coiled wire carrying an electric current produces a magnetic field with characteristics similar to those of a bar magnet. Magnetic field strength and field density depend on the magnitude of the current and the number of coil turns. Note: the flow of positive current is opposite that of the electron flow.

conceptualized as the number of magnetic lines of force per unit area. The magnetic field drops off with the square of the distance. The SI unit for B is the tesla (T), and as a benchmark, the earth's magnetic field is about $1/20,000$ T. An alternate unit is the gauss (G), where $1 \text{ T} = 10,000 \text{ G}$.

Magnetic fields can be induced by a moving charge in a wire (for example, see the section on transformers in Chapter 5). The direction of the magnetic field depends on the sign and the direction of the charge in the wire, as described by the "right hand rule": The fingers point in the direction of the magnetic field when the thumb points in the direction of a moving positive charge (i.e., opposite the direction of electron movement). Wrapping the current-carrying wire many times in a coil causes a superimposition of the magnetic fields, augmenting the overall strength of the magnetic field inside the coil, with a rapid falloff of field strength outside the coil (see Fig. 14-1B). Direct current (DC) amplitude in the coil determines the overall magnitude of the magnetic field strength. In essence, this is the basic design of the "air core" magnets used for diagnostic imaging, which have magnetic field strengths ranging from 0.3 to 2.0 T, where the strength is directly related to the current.

Magnetic Characteristics of the Nucleus

The nucleus exhibits magnetic characteristics on a much smaller scale as described in the previous section for atoms and molecules and their associated electron distributions. The nucleus is comprised of protons and neutrons with characteristics listed in Table 14-1. Magnetic properties are influenced by the spin and charge distributions intrinsic to the proton and neutron. For the proton, which has a unit positive charge (equal to the electron charge but of opposite sign), the nuclear "spin" produces a magnetic dipole. Even though the neutron is electrically uncharged, charge inhomogeneities on the subnuclear scale result in a magnetic field of opposite direction and of approximately the same strength as the proton. The *magnetic moment*, represented as a vector indicating magnitude and direction, describes the magnetic field characteristics of the nucleus. A phenomenon known as *pairing* occurs within the nucleus of the atom, where the constituent protons and neutrons determine the nuclear magnetic moment. If the total number of protons (P) and neutrons (N) in the nucleus is even, the magnetic moment is essentially zero. However, if N is even and P is odd, or N is odd and P is even, the noninteger nuclear spin generates a magnetic moment. A single atom does not generate a large enough nuclear magnetic moment to be observable; the signal measured by an MRI system is the conglomerate signal of billions of atoms.

TABLE 14-1. PROPERTIES OF THE NEUTRON AND PROTON

Characteristic	Neutron	Proton
Mass (kg)	1.674×10^{-27}	1.672×10^{-27}
Charge (coulomb)	0	1.602×10^{-19}
Spin quantum number	$1/2$	$1/2$
Magnetic moment (joule/tesla)	-9.66×10^{-27}	1.41×10^{-26}
Magnetic moment (nuclear magneton)	-1.91	2.79

Nuclear Magnetic Characteristics of the Elements

Biologically relevant elements that are candidates for producing MR images are listed in Table 14-2. The key features include the strength of the magnetic moment, the physiologic concentration, and the isotopic abundance. Hydrogen, having the largest magnetic moment and greatest abundance, is by far the best element for general clinical utility. Other elements are orders of magnitude less sensitive when the magnetic moment and the physiologic concentration are considered together. Of these, ^{23}Na and ^{31}P have been used for imaging in limited situations, despite their relatively low sensitivity. Therefore, the proton is the principal element used for MR imaging.

The spinning proton or "*spin*" (spin and proton are synonymous herein) is classically considered to be like a bar magnet with north and south poles (Fig. 14-2); however, the magnetic moment of a single proton is extremely small and not detectable. A vector representation (amplitude and direction) is helpful when contemplating the additive effects of many protons. Thermal energy agitates and randomizes the direction of the spins in the tissue sample, and as a result there is no net tissue magnetization (Fig. 14-3A). Under the influence of a strong external magnetic field, B_0 , however, the spins are distributed into two energy states: alignment with (parallel to) the applied field at a low-energy level, and alignment against (antiparallel to) the field at a slightly higher energy level (see Fig. 14-3B). A slight majority of spins exist in the low-energy state, the number of which is determined by the thermal energy of the sample (at absolute zero, 0 degrees Kelvin (K), all protons would be aligned in the low-energy state). For higher applied magnetic field strength, the energy separation of the low and high energy levels is greater, as is the number of excess protons in the low-energy state. The number of excess protons in the low-energy state at 1.0 T is about 3 spins per million (3×10^{-6}) at physiologic temperatures and is proportional to the external magnetic field. Although this does not seem significant, for a typical voxel volume in MRI there are about 10^{21} protons, so there are $3 \times 10^{-6} \times 10^{21}$, or approximately 3×10^{15} , more spins in the low-energy state! This number of protons produces an observable magnetic moment when summed.

In addition to energy separation of the spin states, the protons also experience a torque from the applied magnetic field that causes precession, in much the same

TABLE 14-2. MAGNETIC RESONANCE PROPERTIES OF MEDICALLY USEFUL NUCLEI

Nucleus	Spin Quantum Number	% Isotopic Abundance	Magnetic Moment	Relative Physiologic Concentration*	Relative Sensitivity
^1H	$\frac{1}{2}$	99.98	2.79	100	1
^{16}O	0	99.0	0	50	0
^{17}O	$\frac{5}{2}$	0.04	1.89	50	9×10^{-6}
^{19}F	$\frac{1}{2}$	100	2.63	4×10^{-6}	3×10^{-8}
^{23}Na	$\frac{3}{2}$	100	2.22	8×10^{-2}	1×10^{-4}
^{31}P	$\frac{1}{2}$	100	1.13	7.5×10^{-2}	6×10^{-5}

*Note: all isotopes of a given element.

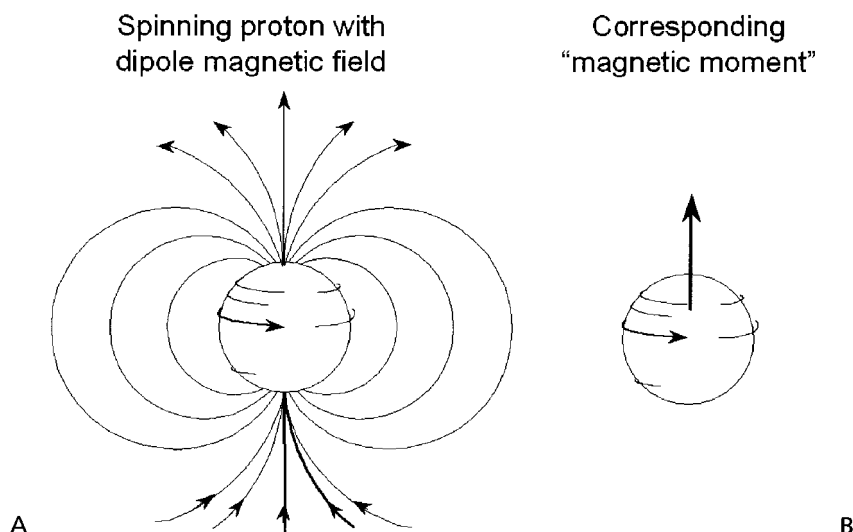


FIGURE 14-2. **A:** A spinning proton can classically be represented with a dipole magnetic field. **B:** Magnetic characteristics of the proton are described as a magnetic moment, with a vector that indicates direction and magnitude.

way that a spinning top wobbles due to the force of gravity (Fig. 14-4A). Direction of the spin axis is perpendicular to the torque's twisting. This precession occurs at an angular frequency (number of rotations/sec about an axis of rotation) that is proportional to the magnetic field strength B_0 . The *Larmor equation* describes the dependence between the magnetic field, B_0 , and the precessional angular frequency, ω_0 :

$$\omega_0 = \gamma B_0$$

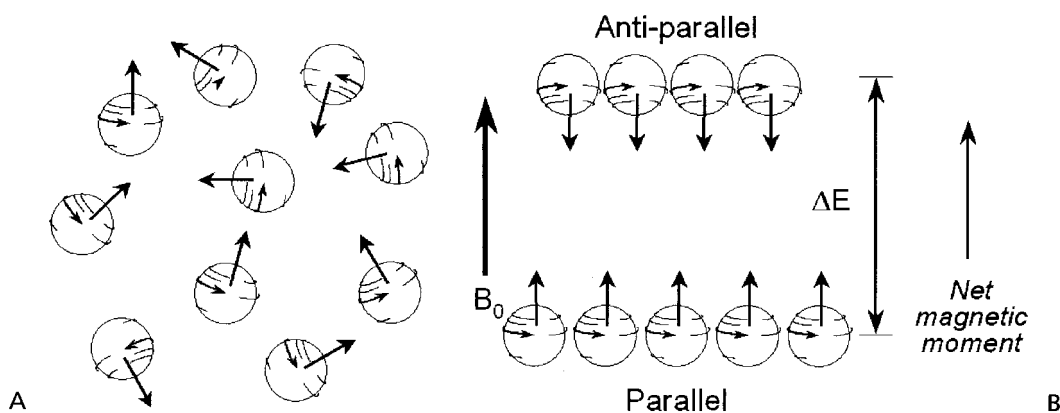


FIGURE 14-3. Simplified distributions of "free" protons without and with an external magnetic field are shown. **A:** Without an external magnetic field, a group of protons assumes a random orientation of magnetic moments, producing an overall magnetic moment of zero. **B:** Under the influence of an applied external magnetic field, B_0 , the protons assume a nonrandom alignment in two possible orientations: parallel and antiparallel to the applied magnetic field. A slightly greater number of protons exist in the parallel direction, resulting in a measurable *sample magnetic moment* in the direction of B_0 .

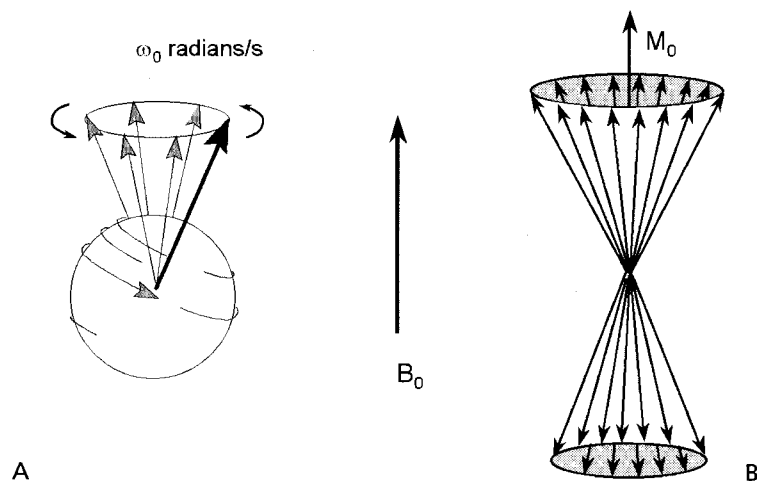


FIGURE 14-4. A: A single proton *precesses* about its axis with an angular frequency, ω , that is proportional to the externally applied magnetic field strength, according to the Larmor equation. **B:** A group of protons in the parallel and antiparallel energy states generates an equilibrium magnetization, M_0 , in the direction of the applied magnetic field B_0 . The protons are distributed randomly over the surface of the cone and produce no magnetization in the perpendicular direction.

or, with respect to linear frequency:

$$f_0 = \frac{\gamma}{2\pi} B_0$$

where γ is the gyromagnetic ratio unique to each element, B_0 is the magnetic field strength in tesla, f is the linear frequency in MHz (where $\omega = 2\pi f$: linear and angular frequency are related by a 2π rotation about a circular path), and $\gamma/2\pi$ is the gyromagnetic ratio expressed in MHz/T. Because energy is proportional to frequency, the energy separation, ΔE , between the parallel and antiparallel spins is proportional to the precessional frequency, and larger magnetic fields produce a higher precessional frequency. Each element has a unique gyromagnetic ratio that allows the discrimination of one element from another, based on the precessional frequency in a given magnetic field strength. In other words, the choice of frequency allows the resonance phenomenon to be tuned to a specific element. The gyromagnetic ratios of selected elements are listed in Table 14-3.

The millions of protons precessing in the parallel and antiparallel directions results in a distribution that can be represented by two cones with the net magnetic moment equal to the vector sum of all the protons in the sample in the direction of the applied magnetic field (see Fig. 14-4B). At equilibrium, no magnetic field exists perpendicular to the direction of the external magnetic field because the individual protons precess with a random distribution, which effectively averages out any net magnetic moment. Energy (in the form of a pulse of radiofrequency electromagnetic radiation) at the precessional frequency (related to ΔE) is absorbed and converts spins from the low-energy, parallel direction to the higher-energy, antiparallel

TABLE 14-3. GYROMAGNETIC RATIO FOR USEFUL ELEMENTS IN MAGNETIC RESONANCE

Nucleus	$\gamma/2\pi$ (MHz/T)
^1H	42.58
^{13}C	10.7
^{17}O	5.8
^{19}F	40.0
^{23}Na	11.3
^{31}P	17.2

direction. As the perturbed system goes back to its equilibrium state, the MR signal is produced.

Typical magnetic field strengths for imaging range from 0.1 to 4.0 T (1,000 to 40,000 G). For protons, the precessional frequency is 42.58 MHz in a 1-T magnetic field (i.e., $\gamma/2\pi = 42.58$ MHz/T for ^1H). The frequency increases or decreases linearly with increases or decreases in magnetic field strength, as shown in the example. Accuracy and precision are crucial for the selective excitation of a given nucleus in a magnetic field of known strength. Spin precession frequency must be known to an extremely small fraction (10^{-12}) of the precessional frequency for modern imaging systems.

Example: What is the frequency of precession of ^1H and ^{31}P at 0.15 T? 0.5 T? 1.5 T? 3.0 T?

The Larmor frequency is calculated as $f_0 = (\gamma/2\pi)B_0$.

Field Strength	0.15 T	0.5 T	1.5 T	3.0 T
^1H	$f = 42.58 \text{ MHz/T} \times 0.15 \text{ T}$ $= 6.39 \text{ MHz}$	42.58×0.5 $= 21.29 \text{ MHz}$	42.58×1.5 $= 63.87 \text{ MHz}$	42.58×3.0 $= 127.74 \text{ MHz}$
^{31}P	$f = 17.2 \text{ MHz/T} \times 0.15 \text{ T}$ $= 2.58 \text{ MHz}$	17.2×0.5 $= 8.6 \text{ MHz}$	17.2×1.5 $= 25.8 \text{ MHz}$	17.2×3 $= 51.6 \text{ MHz}$

The differences in the precessional frequency allow the selective excitation of one elemental species for a given magnetic field strength.

Geometric Orientation

By convention, the applied magnetic field B_0 is directed parallel to the z-axis of the three-dimensional Cartesian coordinate axis system. The x and y axes are perpendicular to the z direction. For convenience, two frames of reference are used: the *laboratory frame* and the *rotating frame*. The laboratory frame (Fig. 14-5A) is a stationary reference frame from the observer's point of view. The proton's magnetic moment precesses about the z-axis in a circular geometry about the x-y plane. The rotating frame (see Fig. 14-5B) is a *spinning* axis system whereby the angular frequency is equal to the precessional frequency of the protons. In this frame, the spins

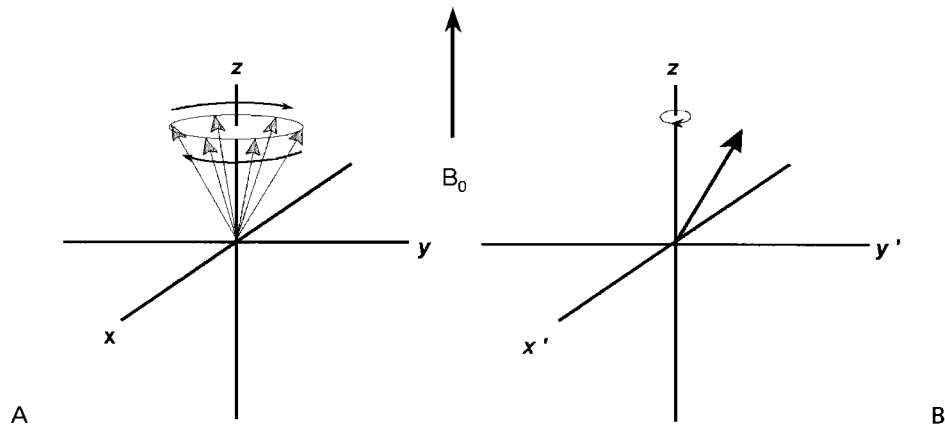


FIGURE 14-5. A: The *laboratory frame of reference* uses stationary three-dimensional Cartesian coordinates. The magnetic moment precesses around the *z*-axis at the Larmor frequency. **B:** The *rotating frame of reference* uses Cartesian coordinate axes that rotate about the *z*-axis at the Larmor precessional frequency, and the other axes are denoted *x'* and *y'*. When precessing at the Larmor frequency, the sample magnetic moment appears stationary.

appear to be stationary when they rotate at the precessional frequency. If a slightly higher precessional frequency occurs, a slow clockwise rotation is observed. For a slightly lower precessional frequency, counterclockwise rotation is observed. A merry-go-round exemplifies an analogy to the laboratory and rotating frames of reference. Externally, from the laboratory frame of reference, the merry-go-round rotates at a specific angular frequency (e.g., 15 rotations per minute [rpm]). Individuals riding the horses are observed moving in a circular path around the axis of rotation, and up-and-down on the horses. If the observer jumps *onto* the merry-go-round, everyone on the ride now appears stationary (with the exception of the up-and-down motion of the horses)—this is the *rotating* frame of reference. Even though the horses are moving up and down, the ability to study them in the rotating frame is significantly improved compared with the laboratory frame. If the merry-go-round consists of three concentric rings that rotate at 14, 15, and 16 rpm and the observer is on the 15-rpm section, all individuals on that particular ring would appear stationary, but individuals on the 14-rpm ring would appear to be rotating in one direction at a rate of 1 rpm, and individuals on the 16-rpm ring would appear to be rotating in the other direction at a rate of 1 rpm. Both the laboratory and the rotating frame of reference are useful in explaining various interactions of the protons with externally applied static and rotating magnetic fields.

The net magnetization vector, M , is described by three components. M_z is the component of the magnetic moment parallel to the applied magnetic field and is known as *longitudinal magnetization*. At equilibrium, the longitudinal magnetization is maximal and is denoted as M_0 , the equilibrium magnetization, where $M_0 = M_z$, with the amplitude determined by the excess number of protons that are in the low-energy state (i.e., aligned with B_0). M_{xy} is the component of the magnetic moment perpendicular to the applied magnetic field and is known as *transverse magnetization*. At equilibrium, the transverse magnetization is zero, because the vector components of the spins are randomly oriented about 360 degrees in the *x*-*y*

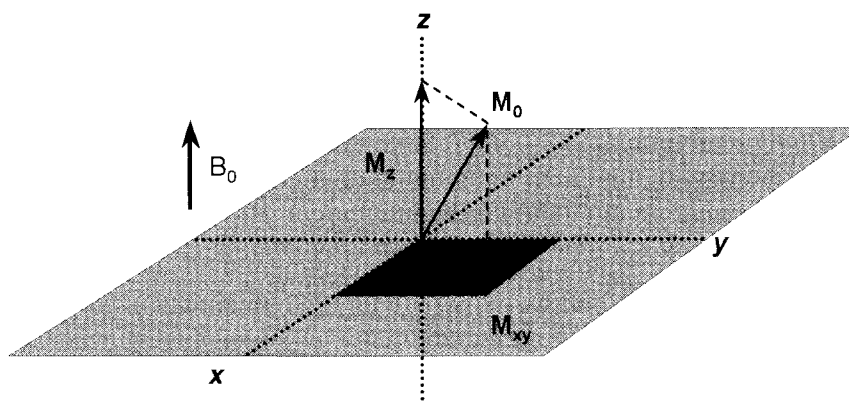


FIGURE 14-6. Longitudinal magnetization, M_z , is the vector component of the magnetic moment in the z direction. Transverse magnetization, M_{xy} , is the vector component of the magnetic moment in the x - y plane. Equilibrium magnetization, M_0 , is the maximal longitudinal magnetization of the sample; in this illustration it is shown displaced from the z -axis.

plane and cancel each other. When the system absorbs energy, M_z is “tipped” into the transverse plane; why this is important is explained in the next section. Figure 14-6 illustrates these concepts.

The famous *Bloch equations* are a set of coupled differential equations that describe the behavior of the magnetization vectors under any conditions. These equations, when integrated, yield the x' , y' , and z components of magnetization as a function of time.

14.2 GENERATION AND DETECTION OF THE MAGNETIC RESONANCE SIGNAL

Application of radiofrequency (RF) energy synchronized to the precessional frequency of the protons causes displacement of the tissue magnetic moment from equilibrium conditions (i.e., more protons are in the antiparallel orientation). Return to equilibrium results in emission of MR signals proportional to the number of excited protons in the sample, with a rate that depends on the characteristics of the tissues. Excitation, detection, and acquisition of the signals constitute the basic information necessary for MR spectroscopy and imaging.

Resonance and Excitation

The displacement of the equilibrium magnetization occurs when the magnetic component of the RF pulse, also known as the B_1 field, is precisely matched to the precessional frequency of the protons to produce a condition of *resonance*. This RF frequency (proportional to energy) corresponds to the energy separation between the protons in the parallel and antiparallel directions, as described by either a quantum mechanics or a classical physics approach. Each has its advantages in the understanding of MR physics.

The quantum mechanics approach considers the RF energy as photons (quanta) instead of waves. Spins oriented parallel and antiparallel to the external

magnetic field are separated by an energy gap, ΔE . Only when the *exact* energy is applied do the spins flip (i.e., transition from the low- to the high-energy level or from the high- to the low-energy level). This corresponds to a *specific frequency* of the RF pulse, equal to the precessional frequency of the spins. The amplitude and duration of the RF pulse determine the overall energy absorption and the number of protons that undergo the energy transition. Longitudinal magnetization changes from the maximal positive value at equilibrium, through zero, to the maximal negative value (Fig. 14-7). Continued RF application induces a return to equilibrium conditions, as an incoming “photon” causes the spontaneous emission of two photons and reversion of the proton to the parallel direction. To summarize, the quantum mechanical description explains the exchange of energy occurring between magnetized protons and photons and the corresponding change in the longitudinal magnetization. However, it does not directly explain how the sample magnetic moment induces a current in a coil and produces the MR signal. The classical physics model better explains this phenomenon.

Example: Determine the energy difference, ΔE (in eV), of the parallel and antiparallel spin states under the influence of a 1.5-T magnetic field and compare that value with an x-ray photon of 50 keV. Information needed includes the relationship between energy and wavelength— E (in keV) = $1.24/\lambda$ (in nm); between magnetic field strength and frequency— $\omega_0 = \gamma B_0$; and between frequency and wavelength— $\lambda = c/f$.

Answer: The precessional frequency, $\omega_0 = \gamma B_0 = 42.58 \text{ MHz/T} \times 1.5 \text{ T} = 63.86 \text{ MHz}$; $\lambda = c/f = (3.0 \times 10^8 \text{ m/sec}) / (63.86 \times 10^6 \text{ /sec}) = 4.7 \text{ m}$; $E = 1.24/\lambda = 1.24/(4.69 \times 10^9) = 2.64 \times 10^{-10} \text{ keV} = 2.64 \times 10^{-7} \text{ eV} = \Delta E$. The 50-keV photon therefore possesses 1.9×10^8 more energy than the photon that causes a magnetized proton to flip in a 1.5-T magnetic field.

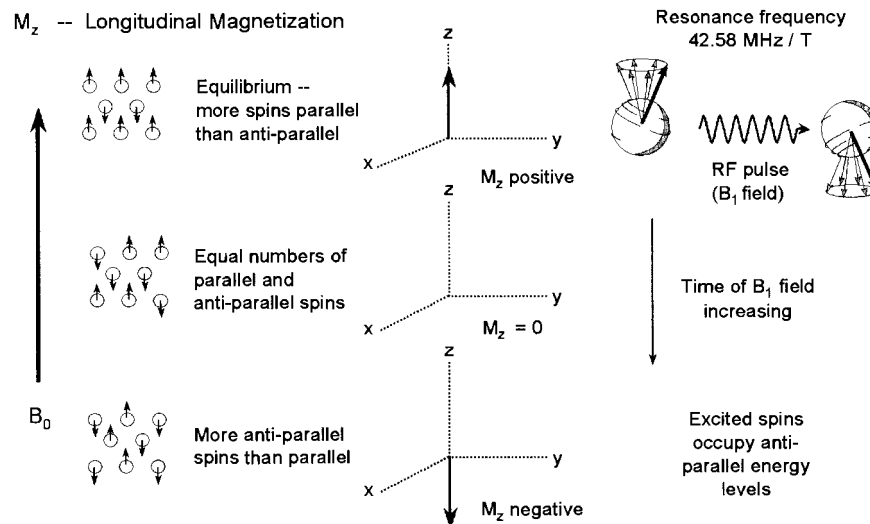


FIGURE 14-7. A simple quantum mechanics process depicts the discrete energy absorption and the time change of the longitudinal magnetization vector as radiofrequency (RF) energy, equal to the energy difference between the parallel and antiparallel spins, is applied to the sample (at the Larmor frequency). Discrete quanta absorption changes the proton energy from parallel to antiparallel. With continued application of RF energy at the Larmor frequency, M_z is displaced from equilibrium, through zero, to the opposite direction (high-energy state).

From the classical physics viewpoint, the B_1 field is considered the magnetic component of an electromagnetic wave with sinusoidally varying electric and magnetic fields. The magnetic field variation can be thought of as comprising two magnetic field vectors of equal magnitude, rotating in opposite directions around a point at the Larmor frequency and traveling at the speed of light (Fig. 14-8). At 0, 180, and 360 degrees about the circle of rotation, the vectors cancel, producing no magnetic field. At 90 degrees, the vectors positively add and produce the peak magnetic field, and at 270 degrees the vectors negatively add and produce the peak magnetic field in the opposite direction, thus demonstrating the characteristics of the magnetic variation. Now consider the magnetic vectors independently; of the two, only one vector will be rotating in the same direction as the precessing spins in the magnetized sample (the other vector will be rotating in the opposite direction). From the perspective of the *rotating frame*, this magnetic vector (the applied B_1 field) is stationary relative to the precessing protons within the sample, and it is applied along the x' -axis (or the y' -axis) in a direction *perpendicular* to the sample magnetic moment, M_z (Fig. 14-9A). The stationary B_1 field applies a torque to M_z , causing a rotation away from the longitudinal direction into the transverse plane. The rotation of M_z occurs at an angular frequency equal to $\omega_1 = \gamma B_1$. Because the tip angle, θ , is equal to $\omega_1 \times t$, where t is the time of the B_1 field application, then, by substitution, $\theta = \gamma B_1 t$, which shows that the time of the applied B_1 field determines the amount of rotation of M_z . Applying the B_1 field in the opposite direction (180-degree change in phase) changes the tip direction of the sample moment. If the RF energy is not applied at the precessional (Larmor) frequency, the B_1 field is not stationary in the rotating frame and the coupling (resonance) between the two magnetic fields does not occur (see Fig. 14-9B).

Flip angles describe the rotation through which the longitudinal magnetization vector is displaced to generate the transverse magnetization (Fig. 14-10). Common angles are 90 degrees ($\pi/2$) and 180 degrees (π), although a variety of smaller (less than 90 degrees) and larger angles are chosen to enhance tissue contrast in various

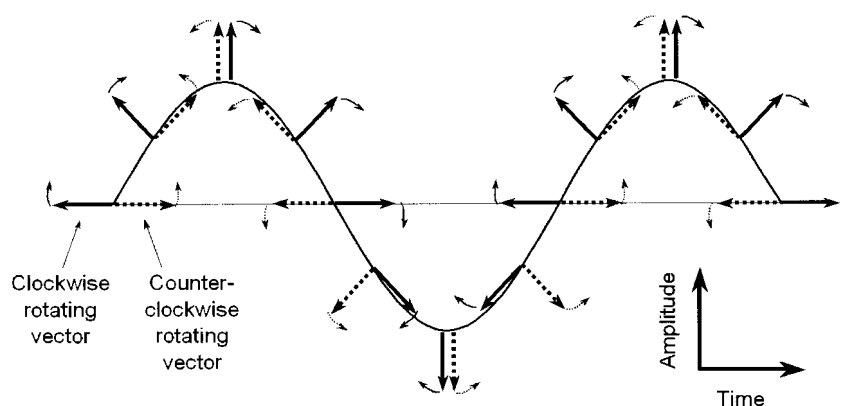


FIGURE 14-8. A classical physics description of the magnetic field component of the radiofrequency pulse (the electric field is not shown). Clockwise (*solid*) and counter-clockwise (*dotted*) rotating magnetic vectors produce the magnetic field variation by constructive and destructive interaction. At the Larmor frequency, one of the magnetic field vectors rotates synchronously in the rotating frame and is therefore stationary (the other vector rotates in the opposite direction).

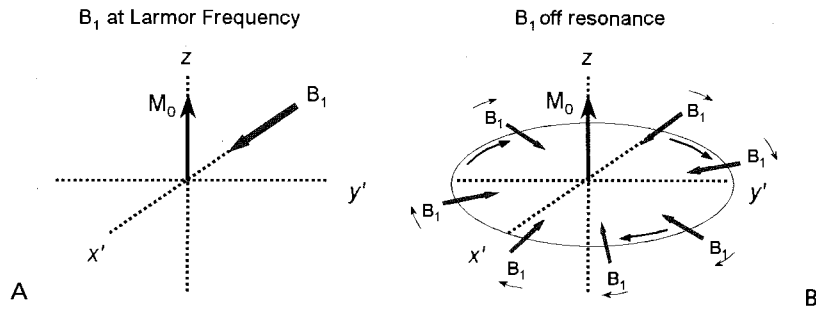


FIGURE 14-9. A: In the rotating frame, the radiofrequency pulse (B_1 field) is applied at the Larmor frequency and is stationary in the $x'-y'$ plane. The B_1 field interacts at 90 degrees to the sample magnetic moment and produces a torque that displaces the magnetic vector away from equilibrium. **B:** The B_1 field is not tuned to the Larmor frequency and is not stationary in the rotating frame. No interaction with the sample magnetic moment occurs.

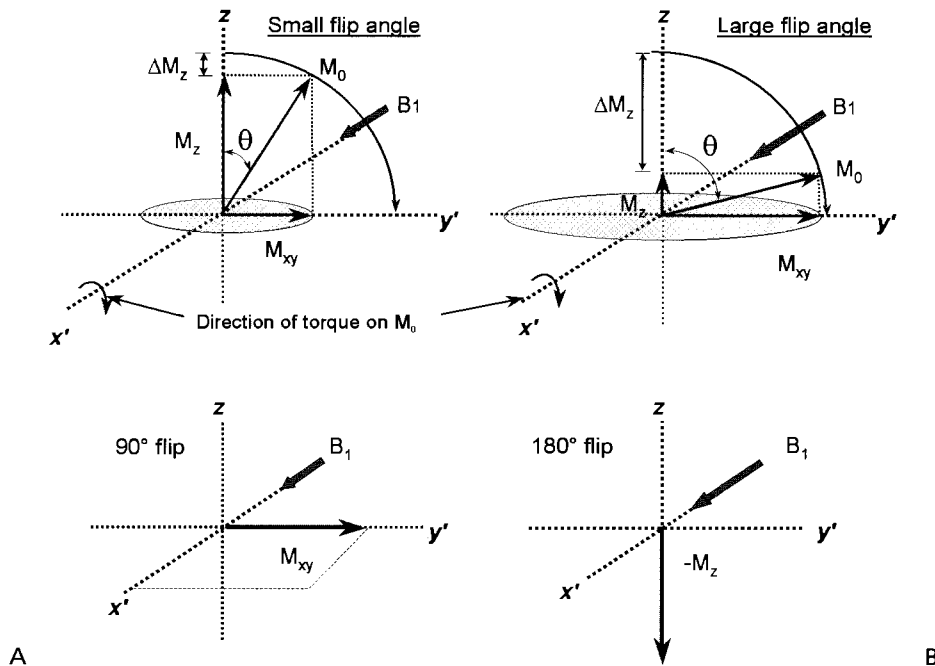


FIGURE 14-10. A: Flip angles describe the angular displacement of the longitudinal magnetization vector from the equilibrium position. The rotation angle of the magnetic moment vector depends on the duration and amplitude of the B_1 field at the Larmor frequency. Small flip angles (~30 degrees) and large flip angles (~90 degrees) produce small and large transverse magnetization, respectively. **B:** Common flip angles are 90 degrees, which produces the maximum transverse magnetization, and 180 degrees, which inverts the existing longitudinal magnetization M_z to $-M_z$.

ways. A 90-degree angle provides the largest possible transverse magnetization. The time required to flip the magnetic moment is linearly related to the displacement angle: For the same B_1 field strength, a 90-degree angle takes half the time to produce that a 180-degree angle does. The time required to implement a rotation is on the order of tens to hundreds of microseconds. With fast MR imaging techniques, 30-degree and smaller angles are often used to reduce the time needed to displace the longitudinal magnetization and generate the transverse magnetization. For flip angles smaller than 90 degrees, less signal in the M_{xy} direction is generated, but less time is needed to displace M_z , resulting in a *greater amount of transverse magnetization (signal) per excitation time*. For instance, a 45-degree flip takes half the time of a 90-degree flip yet creates 70% of the signal, because the projection of the vector onto the transverse plane is $\sin 45$ degrees, or 0.707. In instances where short excitation times are necessary, small flip angles are employed.

Free Induction Decay: T2 Relaxation

The 90-degree RF pulse produces phase coherence of the individual protons and generates the maximum possible transverse magnetization for a given sample volume. As M_{xy} rotates at the Larmor frequency, the receiver antenna coil (in the *laboratory* frame) is induced (by *magnetic induction*) to produce a damped sinusoidal electronic signal known as the *free induction decay* (FID) signal, as shown in Fig. 14-11.

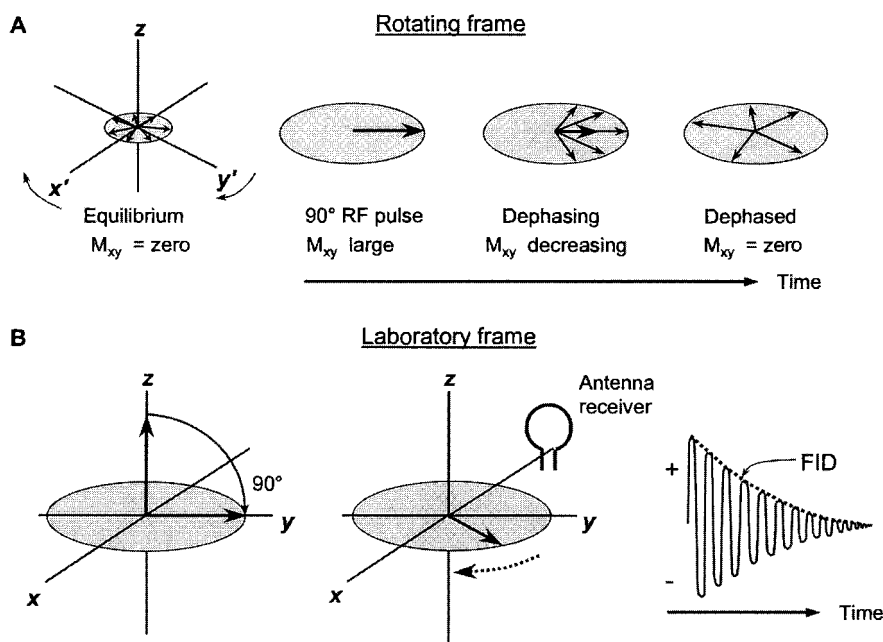


FIGURE 14-11. A: Conversion of longitudinal magnetization, M_z , into transverse magnetization, M_{xy} , results in an initial phase coherence of the individual spins of the sample. The magnetic moment vector precesses at the Larmor frequency (stationary in the rotating frame) and dephases with time. **B:** In the laboratory frame, M_{xy} precesses and induces a signal in an antenna receiver that is sensitive to transverse magnetization. A free induction decay (FID) signal is produced, oscillating at the Larmor frequency, and decays with time due to the loss of phase coherence.

The “decay” of the FID envelope is the result of the loss of phase coherence of the individual spins caused by magnetic field variations. Micromagnetic inhomogeneities intrinsic to the structure of the sample cause a spin-spin interaction, whereby the individual spins precess at different frequencies due to slight changes in the local magnetic field strength. Some spins travel faster and some slower, resulting in a loss of phase coherence. External magnetic field inhomogeneities arising from imperfections in the magnet or disruptions in the field by paramagnetic or ferromagnetic materials accelerate the dephasing process. Exponential relaxation decay, T_2 , represents the intrinsic *spin-spin interactions* that cause loss of phase coherence due to the intrinsic magnetic properties of the sample. The elapsed time between the peak transverse signal and 37% of the peak level ($1/e$) is the T_2 decay constant (Fig. 14-12A). Mathematically, this exponential relationship is expressed as follows:

$$M_{xy}(t) = M_0 e^{-t/T_2}$$

where M_{xy} is the transverse magnetic moment at time t for a sample that has M_0 transverse magnetization at $t = 0$. When $t = T_2$, then $e^{-1} = 0.37$, and $M_{xy} = 0.37 M_0$. An analogous comparison to T_2 decay is that of radioactive decay, with the exception that T_2 is based on $1/e$ decay instead of half-life ($1/2$) decay. This means that the time for the FID to reach half of its original intensity is given by $t = 0.693 \times T_2$.

T_2 decay mechanisms are determined by the molecular structure of the sample. Mobile molecules in amorphous liquids (e.g., cerebral spinal fluid [CSF]) exhibit a long T_2 , because fast and rapid molecular motion reduces or cancels intrinsic magnetic inhomogeneities. As the molecular size increases, constrained molecular motion causes the magnetic field variations to be more readily manifested and T_2 decay to be more rapid. Thus large, nonmoving structures with stationary magnetic inhomogeneities have a very short T_2 .

In the presence of *extrinsic* magnetic inhomogeneities, such as the imperfect main magnetic field, B_0 , the loss of phase coherence occurs more rapidly than from

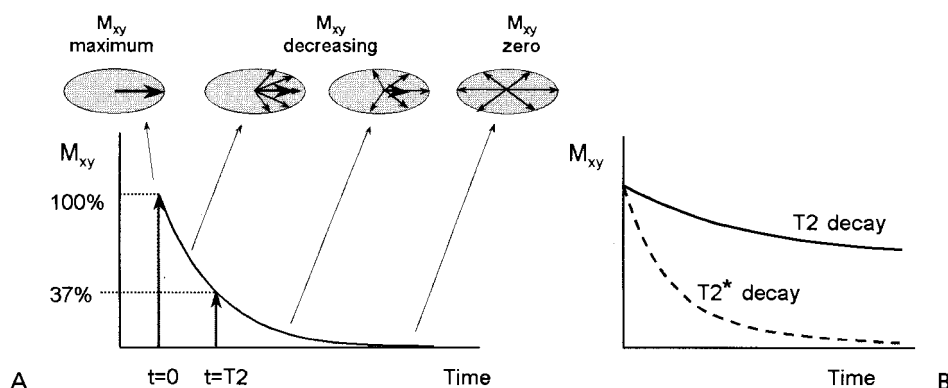


FIGURE 14-12. A: The loss of M_{xy} phase coherence occurs exponentially and is caused by intrinsic spin-spin interactions in the tissues, as well as extrinsic magnetic field inhomogeneities. The exponential decay constant, T_2 , is the time over which the signal decays to 37% of the maximal transverse magnetization (e.g., after a 90-degree pulse). **B:** T_2 is the decay time that results from intrinsic magnetic properties of the sample. T_2^* is the decay time resulting from both intrinsic and extrinsic magnetic field variations. T_2 is always longer than T_2^* .

spin-spin interactions by themselves. When B_0 inhomogeneity is considered, the spin-spin decay constant T_2 is shortened to T_2^* . Figure 14-12B shows a comparison of the T_2 and T_2^* decay curves. T_2^* depends on the homogeneity of the main magnetic field and susceptibility agents that are present in the tissues (e.g., MR contrast materials, paramagnetic or ferromagnetic objects).

Return to Equilibrium: T1 Relaxation

The loss of transverse magnetization (T_2 decay) occurs relatively quickly, whereas the return of the excited magnetization to equilibrium (maximum longitudinal magnetization) takes a longer time. Individual excited spins must release their energy to the local tissue (the lattice). *Spin-lattice relaxation* is a term given for the exponential *regrowth* of M_z , and it depends on the characteristics of the spin interaction with the *lattice* (the molecular arrangement and structure). The T_1 relaxation constant is the time needed to *recover* 63% of the longitudinal magnetization, M_z , after a 90-degree pulse (when $M_z = 0$). The recovery of M_z versus time after the 90-degree RF pulse is expressed mathematically as follows:

$$M_z(t) = M_0(1 - e^{-t/T_1})$$

where M_z is the longitudinal magnetization that recovers after a time t in a material with a relaxation constant T_1 . Figure 14-13 illustrates the recovery of M_z . When $t = T_1$, then $1 - e^{-1} = 0.63$, and $M_z = 0.63 M_0$. Full longitudinal recovery depends on the T_1 time constant. For instance, at a time equal to $3 \times T_1$ after a 90-degree pulse, 95% of the equilibrium magnetization is reestablished. After a period of $5 \times T_1$, the sample is considered to be back to full longitudinal magnetization. A

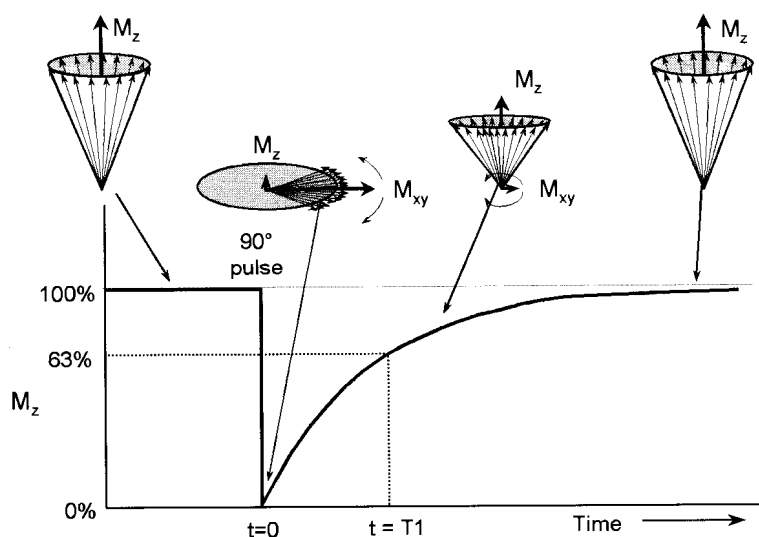


FIGURE 14-13. After a 90-degree pulse, longitudinal magnetization (M_z) is converted from a maximum value at equilibrium to zero. Return of M_z to equilibrium occurs exponentially and is characterized by the spin-lattice T_1 relaxation constant. After an elapsed time equal to T_1 , 63% of the longitudinal magnetization is recovered. Spin-lattice recovery takes longer than spin-spin decay (T_2).

method to determine the T_1 time of a specific tissue or material is illustrated in Fig. 14-14. An initial 90-degree pulse, which takes the longitudinal magnetization to zero, is followed by a delay time, ΔT , and then a second 90-degree pulse is applied to examine the M_z recovery by displacement into the transverse plane (only M_{xy} magnetization can be directly measured). By repeating the sequence with different delay times, ΔT , between 90-degree pulses (from equilibrium conditions), data points that lie on the T_1 recovery curve are determined. The T_1 value can be estimated from these values.

T_1 relaxation depends on the dissipation of absorbed energy into the surrounding molecular lattice. The relaxation time varies substantially for different tissue structures and pathologies. From a classical physics perspective, energy transfer is most efficient when the precessional frequency of the excited protons overlaps with the "vibrational" frequencies of the molecular lattice. Large, slowly moving molecules exhibit low vibrational frequencies that concentrate in the lowest part of the frequency spectrum. Moderately sized molecules (e.g., proteins) and viscous fluids produce vibrations across an intermediate frequency range. Small molecules have vibrational frequencies with low-, intermediate-, and high-frequency components that span the widest frequency range. Therefore, T_1 relaxation is strongly dependent on the physical characteristics of the tissues, as illustrated in Fig. 14-15. Consequently, for *solid and slowly moving structures*, low-frequency variations exist and there is little spectral overlap with the Larmor frequency. A small spectral overlap also occurs for unstructured tissues and fluids that exhibit a wide vibrational frequency spectrum but with low amplitude. In either situation, the inability to release

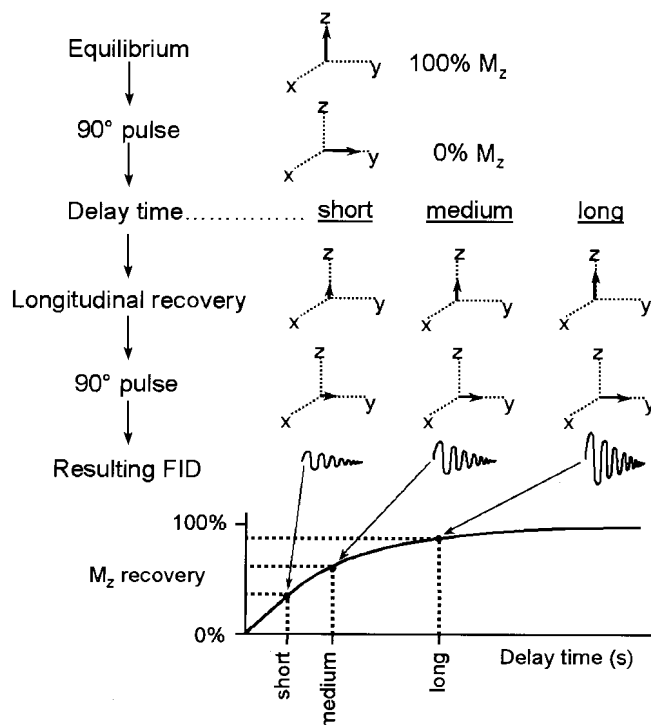


FIGURE 14-14. Spin-lattice relaxation for a sample can be measured by using various delay times between two 90-degree radiofrequency pulses. After an initial 90-degree pulse, longitudinal magnetization (M_z) is equal to zero; after a known delay, another 90-degree pulse is applied, and the longitudinal magnetization that has recovered during the delay is converted to transverse magnetization. The maximum amplitude of the resultant free induction decay (FID) is recorded as a function of delay time, and the points are fit to an exponential recovery function to determine T_1 .

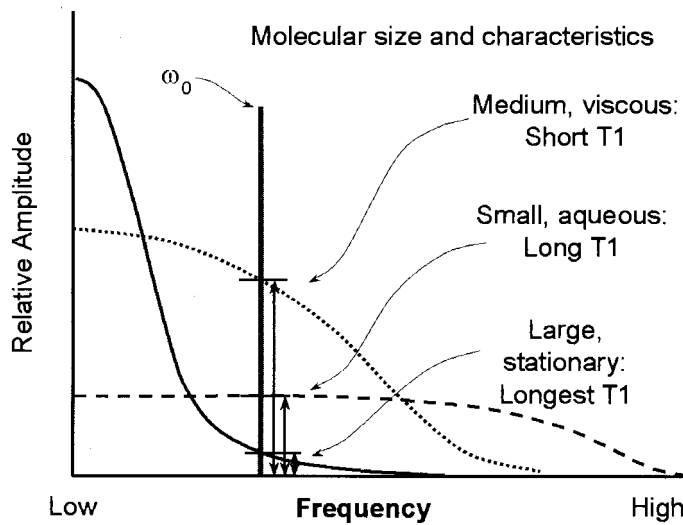


FIGURE 14-15. A classical physics explanation of spin-lattice relaxation is based on the vibration of the molecular lattice of a sample material and its frequency spectrum. Large, stationary structures exhibit little motion with mostly low-frequency vibrations (**solid curve**). Medium sized, proteinated materials have increased frequency amplitudes (**short dash curve**), and small sized, aqueous materials have frequencies distributed over a broad range (**long dash curve**). The overlap of the Larmor precessional frequency (**vertical bar**) with the molecular vibration spectrum indicates the probability of spin-lattice relaxation.

energy to the lattice results in a relatively long T1 relaxation. One interesting case is that of water, which has an extremely long T1, but the addition of water-soluble proteins produces a hydration layer that slows the molecular vibrations and shifts the high frequencies in the spectrum to lower values that increase the amount of spectral overlap with the Larmor frequency and result in a dramatically shorter T1. Moderately sized molecules, such as lipids, proteins, and fats, have a more structured lattice with a vibrational frequency spectrum that is most conducive to spin-lattice relaxation. For biologic tissues, T1 ranges from 0.1 to 1 second in soft tissues, and from 1 to 4 seconds in aqueous tissues (e.g., CSF) and water.

T1 relaxation increases with higher field strengths. A corresponding increase in the Larmor precessional frequency reduces the spectral overlap of the molecular vibrational frequency spectrum, resulting in longer T1 times. Contrast agents (e.g., complex macromolecules containing gadolinium) are effective in decreasing T1 relaxation time by allowing free protons to become bound and create a hydration layer, thus providing a spin-lattice energy sink and a rapid return to equilibrium. Even a very small amount of gadolinium contrast in pure water has a dramatic effect on T1, decreasing the relaxation from a couple of seconds to tens of milliseconds!

Comparison of T1 and T2

T1 is significantly longer than T2. For instance, in a soft tissue, a T1 time of 500 msec has a corresponding T2 time that is typically 5 to 10 times shorter (i.e., about

50 msec). Molecular motion, size, and interactions influence T1 and T2 relaxation (Fig. 14-16). Molecules can be categorized roughly into three size groups—small, medium, and large—with corresponding fast, medium, and slow vibrational frequencies. For reasons described in the previous two sections, small molecules exhibit long T1 and long T2, and intermediate-sized molecules have short T1 and short T2; however, large, slowly moving or bound molecules have long T1 and short T2 relaxation times. Because most tissues of interest in MR imaging consist of intermediate to small-sized molecules, a long T1 usually infers a long T2, and a short T1 infers a short T2. It is the differences in T1, T2, and T2* (along with proton density variations and blood flow) that provide the extremely high contrast in MRI.

Magnetic field strength influences T1 relaxation but has an insignificant impact on T2 decay. This is related to the dependence of the Larmor frequency on magnetic field strength and the degree of overlap with the molecular vibration spectrum. A higher magnetic field strength increases the Larmor frequency ($\omega_0 = \gamma B_0$), which reduces the amount of spectral overlap and produces a longer T1. In Table 14-4, a comparison of T1 and T2 for various tissues is listed. Agents that disrupt the local magnetic field environment, such as paramagnetic blood degradation products, elements with unpaired electron spins (e.g., gadolinium), or any ferromagnetic materials, cause a significant decrease in T2*. In situations where a macromolecule binds free water into a hydration layer, T1 is also significantly decreased.

To summarize, $T1 > T2 > T2^*$, and the specific relaxation times are a function of the tissue characteristics. The spin density, T1, and T2 decay constants are fundamental properties of tissues, and therefore these tissue properties can be exploited by MRI to aid in the diagnosis of pathologic conditions such as cancer, multiple sclerosis, or hematoma. It is important to keep in mind that T1 and T2 (along with spin density) are fundamental properties of tissue, whereas the other time-dependent parameters are machine-dependent. Mechanisms to exploit the differences in T1, T2, and spin density (proton density) are detailed in the next section.

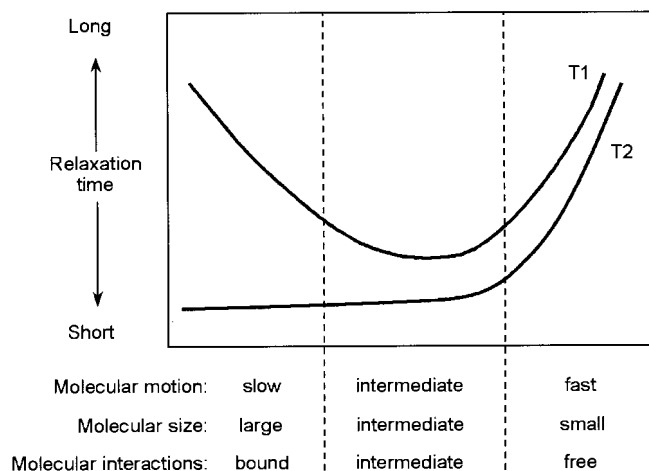


FIGURE 14-16. Factors affecting T1 and T2 relaxation times of different tissues are generally based on molecular motion, size, and interactions. The relaxation times (vertical axis) are different for T1 and T2.

TABLE 14-4. T1 AND T2 RELAXATION CONSTANTS FOR SEVERAL TISSUES^a

Tissue	T1, 0.5 T (msec)	T1, 1.5 T (msec)	T2 (msec)
Fat	210	260	80
Liver	350	500	40
Muscle	550	870	45
White matter	500	780	90
Gray matter	650	900	100
Cerebrospinal fluid	1,800	2,400	160

^aEstimates only, as reported values for T1 and T2 span a wide range.

14.3 PULSE SEQUENCES

Emphasizing the differences among spin density, T1, and T2 relaxation time constants of the tissues is the key to the exquisite contrast sensitivity of MR images. Tailoring the pulse sequences—that is, the timing, order, polarity, and repetition frequency of the RF pulses and applied magnetic field gradients—makes the emitted signals dependent on T1, T2 or spin density relaxation characteristics. MR relies on three major pulse sequences: spin echo, inversion recovery, and gradient recalled echo. When these used in conjunction with localization methods (i.e., the ability to spatially encode the signal to produce an image, discussed in Chapter 15), “contrast-weighted” images are obtained. In the following three sections, the salient points and essential considerations of MR contrast generated by these three major pulse sequences are discussed.

14.4 SPIN ECHO

Spin echo describes the excitation of the magnetized protons in a sample with an RF pulse and production of the FID, followed by a second RF pulse to produce an echo. Timing between the RF pulses allows separation of the initial FID and the echo and the ability to adjust tissue contrast.

Time of Echo

An initial 90-degree pulse produces the maximal transverse magnetization, M_{xy} , and places the spins in phase coherence. The signal exponentially decays with T2* relaxation caused by intrinsic and extrinsic magnetic field variations. After a time delay of TE/2, where TE is the time of echo (defined below), a 180-degree RF pulse is applied, which inverts the spin system and induces a rephasing of the transverse magnetization. The spins are rephased and produce a measurable signal at a time equal to the time of echo (TE). This sequence is depicted in the *rotating* frame in Fig. 14-17. The echo reforms in the *opposite* direction from the initial transverse magnetization vector, so the spins experience the *opposite external magnetic field inhomogeneities* and this strategy cancels their effect.

The B_0 inhomogeneity-canceling effect that the spin echo pulse sequence produces has been likened to a foot race on a track. The racers start running at

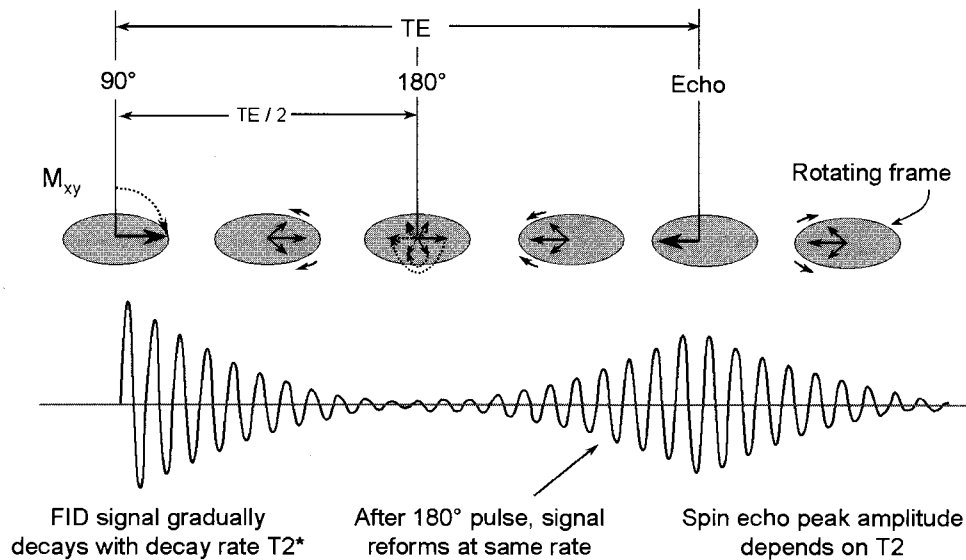


FIGURE 14-17. The spin echo pulse sequence starts with a 90-degree pulse and produces a free induction decay (FID) that decays according to $T2^*$ relaxation. After a delay time equal to $TE/2$, a 180-degree radiofrequency pulse inverts the spins that reestablishes phase coherence and produces an echo at a time TE . Inhomogeneities of external magnetic fields are canceled, and the peak amplitude of the echo is determined by $T2$ decay. The rotating frame of the echo vector in the opposite direction of the FID.

the 90-degree pulse, but quickly their tight grouping at the starting line spreads out (dephases) as they run at different speeds. After a short period, the runners are spread out along the track, with the fastest runners in front and the slower ones in the rear. At this time ($TE/2$), a 180-degree pulse is applied and the runners all instantly reverse their direction, but they keep running at the same speed as before. Immediately after the 180-degree rephasing RF pulse, the fastest runners are the farthest behind and the slowest runners are in front of the pack. Under these conditions, the fast runners at the end of the pack will catch the slow runners at the front of the pack as they all run past the starting line together (i.e., at time TE). Even in a field of runners in which each runs at a markedly different speed from the others, they all will recross the starting line at exactly TE . The MR signal is at a maximum (i.e., the peak of the FID envelope) as the runners are all in phase when they cross the starting line. They can run off in the other direction, and after another time interval of $TE/2$ reverse their direction and run back to the starting line. Again, after a second TE period, they will all cross the starting line (and the FID signal will be at its third peak), then head off in the other direction. This process can be repeated (Fig. 14-18). The maximal echo amplitude depends on the $T2$ constant and not on $T2^*$, which is the decay constant that includes magnetic field inhomogeneities. Of course all MR signals depend on the proton density of the tissue sample, as well. Just before and after the peak amplitude of the echo (centered at time TE), digital sampling and acquisition of the

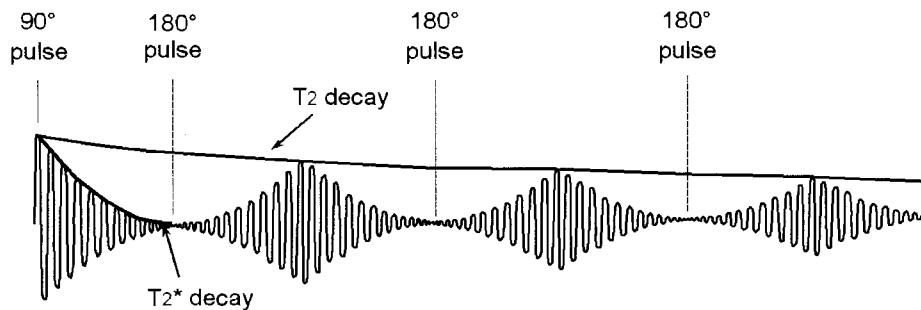


FIGURE 14-18. “True” T2 decay is determined from multiple 180-degree refocusing pulses. While the free induction decay (FID) envelope decays with the T2* decay constant, the peak echo amplitudes of subsequent echoes decay exponentially with time, according to the T2 decay constant.

signal occurs. Spin echo formation separates the RF excitation and signal acquisition events by finite periods of time, which emphasizes the fact that relaxation phenomena are being observed and encoded into the images. Contrast in the image is produced because different tissue types relax differently (based on their T1 and T2 characteristics).

Multiple echoes generated by 180-degree pulses after the initial excitation allow the determination of the “true T2” of the sample. Signal amplitude is measured at several points in time, and an exponential curve is fit to this measured data (see Fig. 14-18). The T2 value is one of the curve-fitting coefficients.

Time of Repetition and Partial Saturation

The standard spin echo pulse sequence uses a series of 90-degree pulses separated by a period known as the *time of repetition* (TR), which typically ranges from about 300 to 3,000 msec. A time delay between excitation pulses allows recovery of the longitudinal magnetization. During this period, the FID and the echo produce the MR signal (explained later). After the TR interval, the next 90-degree pulse is applied, but usually *before* the complete longitudinal magnetization recovery of the tissues. In this instance, the FID generated is less than the first FID. After the second 90-degree pulse, a steady-state longitudinal magnetization produces the same FID amplitude from each subsequent 90-degree pulse (spins are rotated through 360 degrees and are reintroduced in the transverse plane). Tissues become *partially saturated* (i.e., the full transverse magnetization is decreased from the equilibrium magnetization), with the amount of saturation dependent on the T1 relaxation time. A short-T1 tissue has *less* saturation than a long-T1 tissue, as illustrated in Fig. 14-19. For spin echo sequences, partial saturation of the longitudinal magnetization depends on the TR and T1 of the tissues. Partial saturation has an impact on tissue contrast.

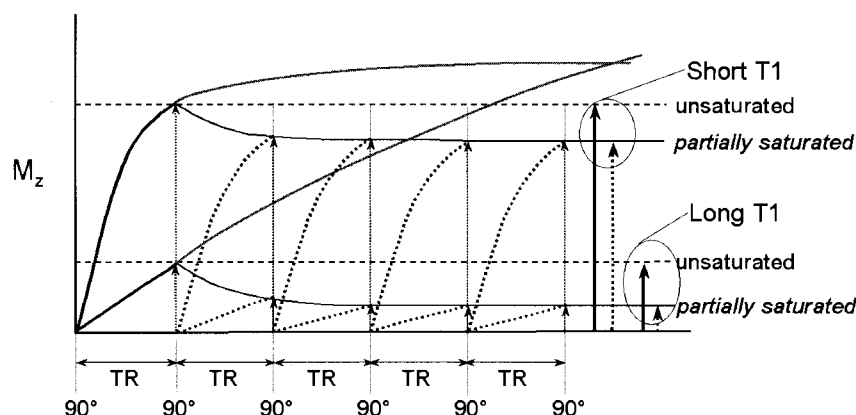


FIGURE 14-19. Partial saturation of tissues occurs because the repetition time between excitation pulses does not completely allow for full return to equilibrium, and the longitudinal magnetization (M_z) amplitude for the next radiofrequency pulse is reduced. After the first excitation pulse, a steady-state equilibrium is reached, wherein the tissues become partially saturated and produce a constant transverse magnetization of less amplitude. Tissues with long T1 experience a greater saturation than do tissues with short T1.

Spin Echo Contrast Weighting

Contrast in an image is proportional to the difference in signal intensity between adjacent pixels in the image, corresponding to two different voxels in the patient. The details of how images are produced spatially in MRI are discussed in Chapter 15. However, the topic of contrast can be discussed here by understanding how the signal changes for different tissue types and for different pulse sequences. The signal, S , produced by an NMR system is proportional to other factors as follows:

$$S \propto \rho_H f(v) [1 - e^{-TR/T1}] e^{-TE/T2}$$

where ρ_H is the spin (proton) density, $f(v)$ is the signal arising from fluid flow (discussed in Chapter 15), T1 and T2 are physical properties of tissue, and TR and TE are pulse sequence controls on the MRI machine (Fig. 14-20).

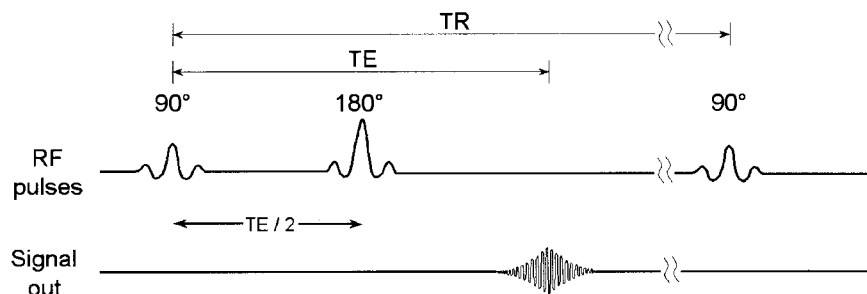


FIGURE 14-20. Spin echo pulse sequence timing is initiated with a 90-degree pulse, followed by a 180-degree pulse applied at time TE/2. The peak echo signal occurs at the echo time, TE. The signal is acquired during the evolution and decay of the echo. After a delay time allowing recovery of M_z , (the end of the TR period) the sequence is repeated many times (e.g., 128 to 256 times) to acquire the information needed to build an image. The repetition time, TR, is the time between 90-degree initiation pulses.

The equation shows that for the same values of TR and TE (i.e., for the same pulse sequence), different values of T1 or T2 (or of ρ_H or $f(v)$) will change the signal S. The signal in adjacent voxels will be different when T1 or T2 changes between those two voxels, and this is the essence of how contrast is formed in MRI. Importantly, by changing the pulse sequence parameters TR and TE, the contrast dependence in the image can be weighted toward T1 or toward T2.

T1 Weighting

A “T1-weighted” spin echo sequence is designed to produce contrast chiefly based on the T1 characteristics of tissues by de-emphasizing T2 contributions. This is achieved with the use of a relatively short TR to maximize the differences in longitudinal magnetization during the return to equilibrium, and a short TE to minimize T2 dependency during signal acquisition. In the longitudinal recovery and transverse decay diagram (Fig. 14-21A), note that the TR time on the abscissa of

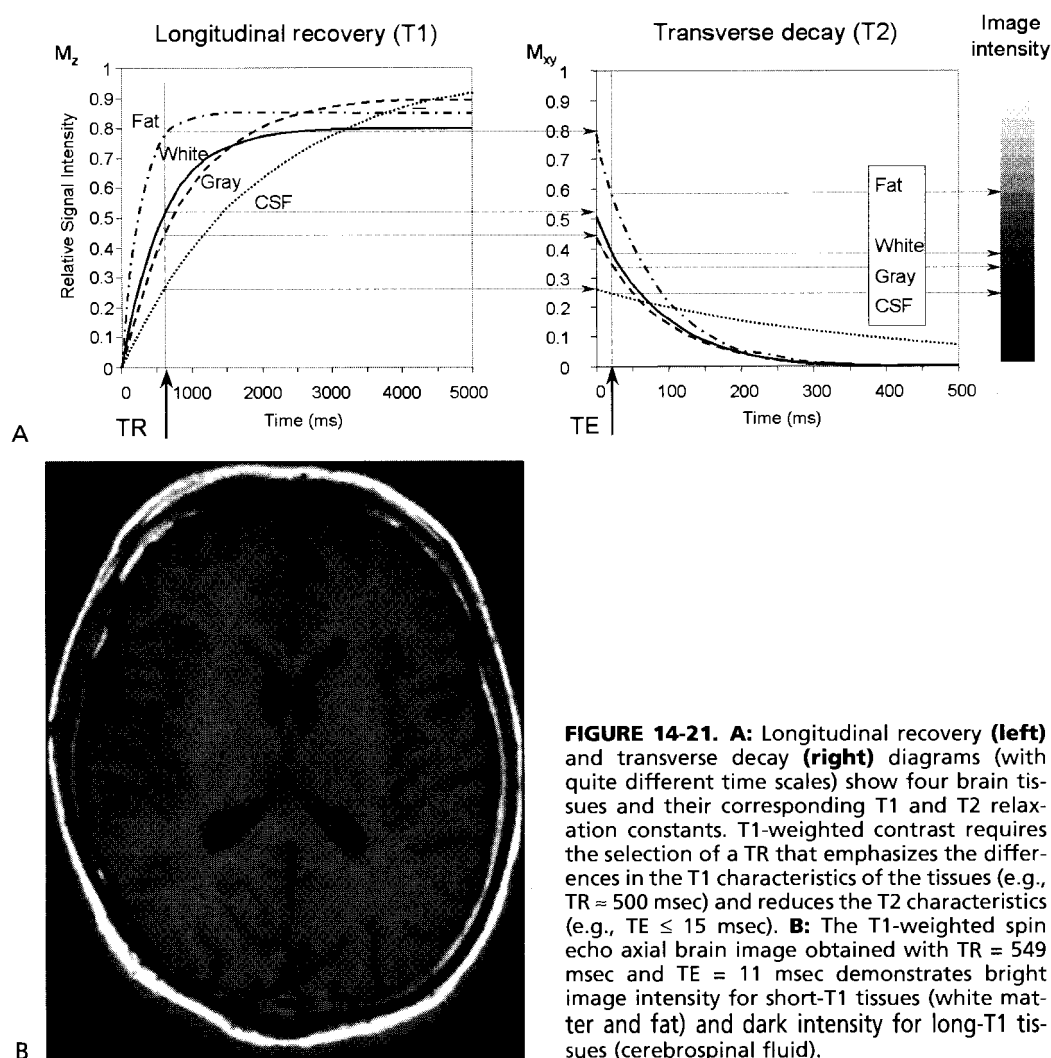


FIGURE 14-21. A: Longitudinal recovery (**left**) and transverse decay (**right**) diagrams (with quite different time scales) show four brain tissues and their corresponding T1 and T2 relaxation constants. T1-weighted contrast requires the selection of a TR that emphasizes the differences in the T1 characteristics of the tissues (e.g., TR = 500 msec) and reduces the T2 characteristics (e.g., TE ≤ 15 msec). **B:** The T1-weighted spin echo axial brain image obtained with TR = 549 msec and TE = 11 msec demonstrates bright image intensity for short-T1 tissues (white matter and fat) and dark intensity for long-T1 tissues (cerebrospinal fluid).

the figure on the left (longitudinal recovery) intersects the individual tissue curves and projects over to the figure on the right (transverse decay). These values represent the amount of magnetization that is available to produce the transverse signal, and therefore the individual tissue curves on right-hand figure start at this point at time $T = 0$. The horizontal projections (*arrows*) graphically demonstrate how the T1 values modulate the overall MRI signal. When TR is chosen to be 400 to 600 msec, the difference in longitudinal magnetization relaxation times (T1) between tissues is emphasized. Four common cerebral tissues—fat, white matter, gray matter, CSF—are shown in the diagrams. The amount of transverse magnetization (which gives rise to a measurable signal) after the 90-degree pulse depends on the amount of longitudinal recovery that has occurred in the tissue of the excited sample. Fat, with a short T1, has a large signal, because the short T1 value allows rapid recovery of the M_z vector. The short T1 value means that the spins rapidly reassume their equilibrium conditions. White and gray matter have intermediate T1 values, and CSF, with a long T1, has a small signal. For the transverse decay (T2) diagram in Fig. 14-21A, a 180-degree RF pulse applied at time TE/2 produces an echo at time TE. A *short* TE preserves the T1 signal differences with minimal transverse decay, which reduces the signal dependence on T2. A long TE is counterproductive in terms of emphasizing T1 contrast, because the signal becomes corrupted with T2 decay. T1-weighted images therefore require a *short TR and a short TE* for the spin echo pulse sequence. A typical T1-weighted axial image of the brain acquired with TR = 500 msec and TE = 8 msec is illustrated in Fig. 14-21B. Fat is the most intense signal (shortest T1); white matter and gray matter have intermediate intensities; and CSF has the lowest intensity (longest T1). A typical spin echo T1-weighted image is acquired with a TR of about 400 to 600 msec and a TE of 5 to 20 msec.

Spin (Proton) Density Weighting

Image contrast with spin density weighting relies mainly on differences in the number of magnetizable protons per volume of tissue. At thermal equilibrium, those tissues with a greater spin density exhibit a larger longitudinal magnetization. Very hydrogenous tissues such as lipids and fats have a high spin density compared with proteinaceous soft tissues; aqueous tissues such as CSF also have a relatively high spin density. Fig. 14-22A illustrates the longitudinal recovery and transverse decay diagram for spin density weighting. To minimize the T1 differences of the tissues, a relatively long TR is used. This allows significant longitudinal recovery so that the transverse magnetization differences are chiefly those resulting from variations in spin density (CSF > fat > gray matter > white matter). Signal amplitude differences in the FID are preserved with a short TE, so the influences of T2 differences are minimized. Spin density-weighted images therefore require a *long TR and a short TE* for the spin echo pulse sequence. Fig. 14-22B shows a spin density-weighted image with TR = 2,400 msec and TE = 30 msec. Fat and CSF display as a relatively bright signal, and a slight contrast inversion between white and gray matter occurs. A typical spin density-weighted image has a TR between 2,000 and 3,500 msec and a TE between 8 and 30 msec. This sequence achieves the highest overall signal and the highest signal-to-noise ratio (SNR) for spin echo imaging; however, the image contrast is relatively poor, and therefore the contrast-to-noise ratio is not necessarily higher than with a T1- or T2-weighted image.

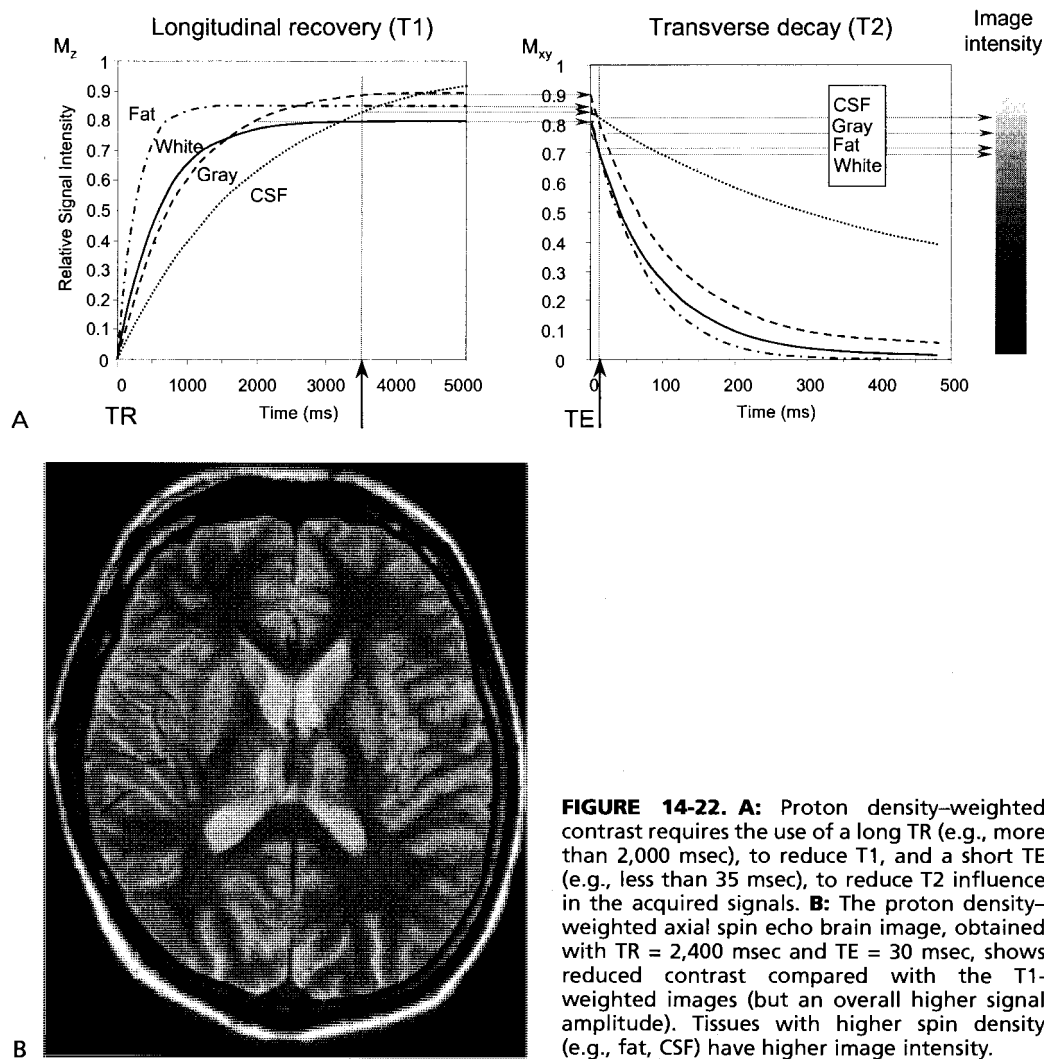


FIGURE 14-22. A: Proton density-weighted contrast requires the use of a long TR (e.g., more than 2,000 msec), to reduce T1, and a short TE (e.g., less than 35 msec), to reduce T2 influence in the acquired signals. **B:** The proton density-weighted axial spin echo brain image, obtained with TR = 2,400 msec and TE = 30 msec, shows reduced contrast compared with the T1-weighted images (but an overall higher signal amplitude). Tissues with higher spin density (e.g., fat, CSF) have higher image intensity.

T2 Weighting

T2 weighting follows directly from the spin density weighting sequence: Reduce T1 effects with a long TR, and accentuate T2 differences with a *longer* TE. The T2-weighted signal is usually the second echo (produced by a second 180-degree pulse) of a long-TR spin echo pulse sequence (the first echo is spin density weighted, with short TE, as explained in the previous section). Generation of T2 contrast differences is shown in Fig. 14-23A. Compared with a T1-weighted image, inversion of tissue contrast occurs (CSF is brighter than fat instead of darker), because short-T1 tissues usually have a short T2, and long-T1 tissues have a long T2. Tissues with a long T2 (e.g., CSF) maintain transverse magnetization longer than short-T2 tissues, and thus result in higher signal intensity. A T2-weighted image (Fig. 14-23B), demonstrates the contrast inversion and high tissue contrast features, compared with the T1-weighted image (Fig. 14-21B). As TE is increased, more T2 contrast is

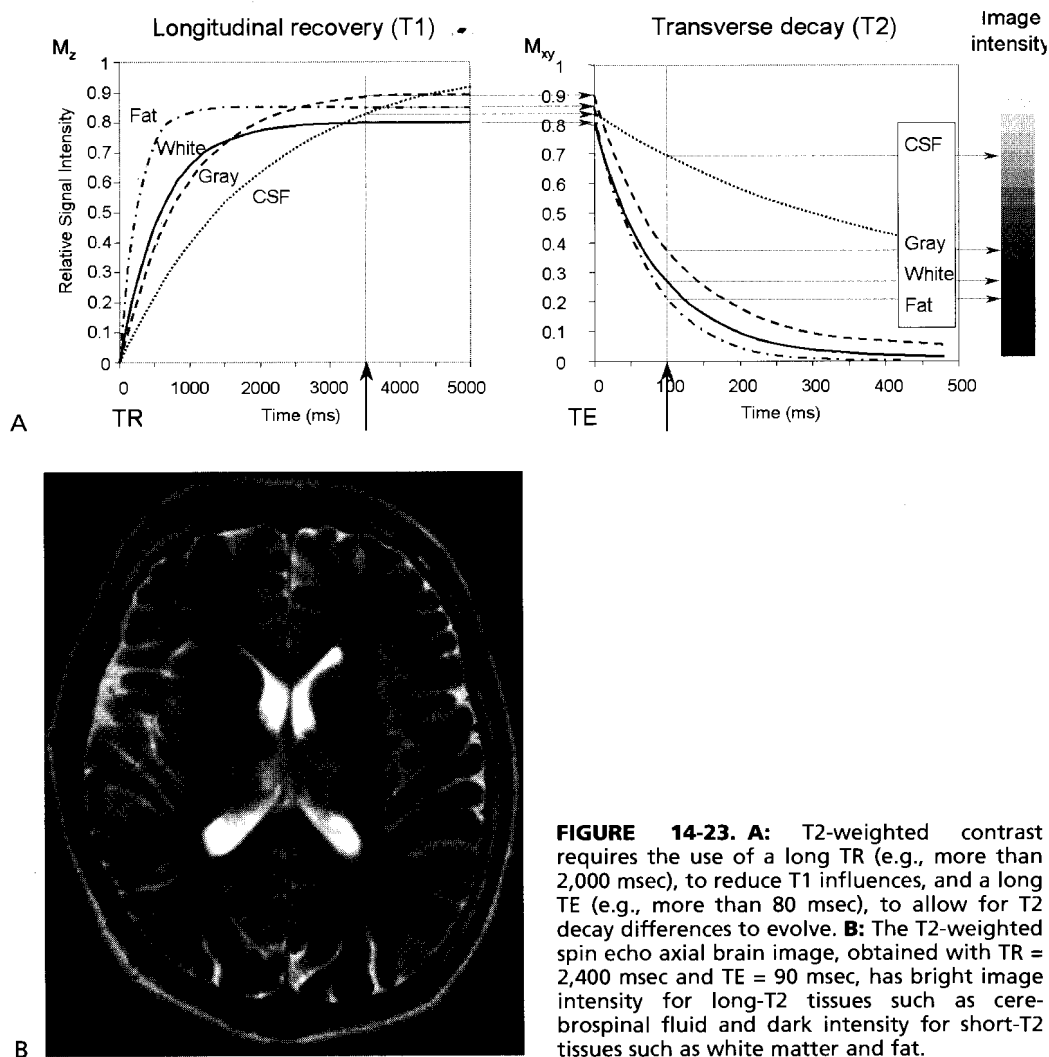


FIGURE 14-23. A: T2-weighted contrast requires the use of a long TR (e.g., more than 2,000 msec), to reduce T1 influences, and a long TE (e.g., more than 80 msec), to allow for T2 decay differences to evolve. **B:** The T2-weighted spin echo axial brain image, obtained with TR = 2,400 msec and TE = 90 msec, has bright image intensity for long-T2 tissues such as cerebrospinal fluid and dark intensity for short-T2 tissues such as white matter and fat.

achieved, at the expense of a reduced transverse magnetization signal. Even with low signal, *window width* and *window level* adjustments (see Chapter 4) remap the signals over the full range of the display, so that the overall perceived intensity is similar for all images. The typical T2-weighted sequence uses a TR of approximately 2,000 to 4,000 msec and a TE of 80 to 120 msec.

Spin Echo Parameters

Table 14-5 lists the typical contrast weighting values of TR and TE for spin echo imaging. For situations intermediate to those listed, mixed-contrast images are the likely outcome. For conventional spin echo sequences, both a spin density and a T2-weighted contrast signal are acquired during each TR by acquiring two echoes with a short TE and a long TE (Fig. 14-24).

TABLE 14-5. SPIN ECHO PULSE SEQUENCE CONTRAST WEIGHTING PARAMETERS

Parameter	T1 Contrast	Spin Density Contrast	T2 Contrast
TR (msec)	400–600	1,500–3,500	1,500–3,500
TE (msec)	5–30	5–30	60–150

TE, time of echo; TR, time of repetition.

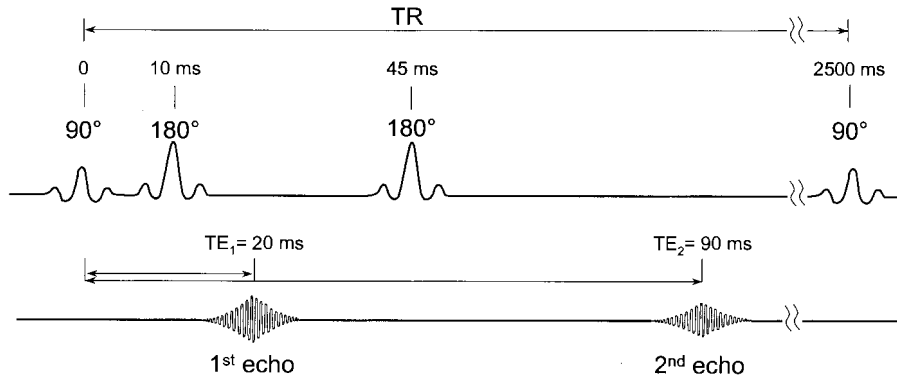


FIGURE 14-24. Multiple echo pulse sequence can produce spin density and T2-weighted contrast from the first and second echoes during the same TR. In this diagram, $TE_1 = 20$ msec and $TE_2 = 90$ msec.

14.5 INVERSION RECOVERY

Inversion recovery, also known as inversion recovery spin echo, emphasizes *T1 relaxation times* of the tissues by extending the amplitude of the longitudinal recovery by a factor of two. An initial 180-degree RF pulse inverts the M_z longitudinal magnetization of the tissues to $-M_z$. After a delay time known as the *time of inversion* (TI), a 90-degree RF pulse rotates the recovered fraction of M_z spins into the transverse plane to generate the FID. A second 180-degree pulse at time $TE/2$ produces an echo signal at TE (Fig. 14-25). *TR in this case is the time between 180-degree initiation pulses.* As with the spin echo pulse sequence, a TR less than about $5 \times T1$ of the tissues causes a partial saturation of the spins and a steady-state equilibrium of the longitudinal magnetization after the first three to four excitations in the sequence. The echo amplitude of a given tissue depends on TI, TE, TR, and the magnitude (positive or negative) of M_z .

The signal intensity at a location (x,y) in the image matrix is approximated as follows:

$$S \propto \rho_H f(v) (1 - 2e^{-TI/T1} + e^{-TR/T1})$$

where $f(v)$ is a function of motion (e.g., blood flow) and the factor of 2 arises from the longitudinal magnetization recovery from $-M_z$ to M_z . Note the lack of T2 dependence in this approximation of S (in fact, there is some T2 decay.) The time

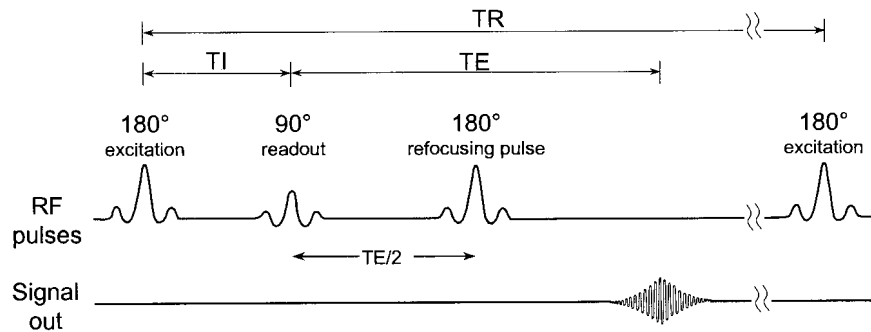


FIGURE 14-25. Inversion recovery spin echo sequence begins with a 180-degree “excitation” pulse to invert the longitudinal magnetization, an inversion time (TI) delay, and then a 90-degree “readout” pulse to convert recovered M_z into M_{xy} . A refocusing 180-degree pulse creates the echo at time TE for signal acquisition, and then the sequence is repeated after a time TR between 180-degree excitation pulses.

of inversion, TI, controls the contrast between tissues (Fig. 14-26). The TE must be kept short to avoid mixed-contrast images and to minimize T2 dependency. TR determines the amount of longitudinal recovery between 180-degree excitation pulses.

The inversion recovery sequence produces “negative” longitudinal magnetization that results in negative (in phase) or positive (out of phase) transverse magne-

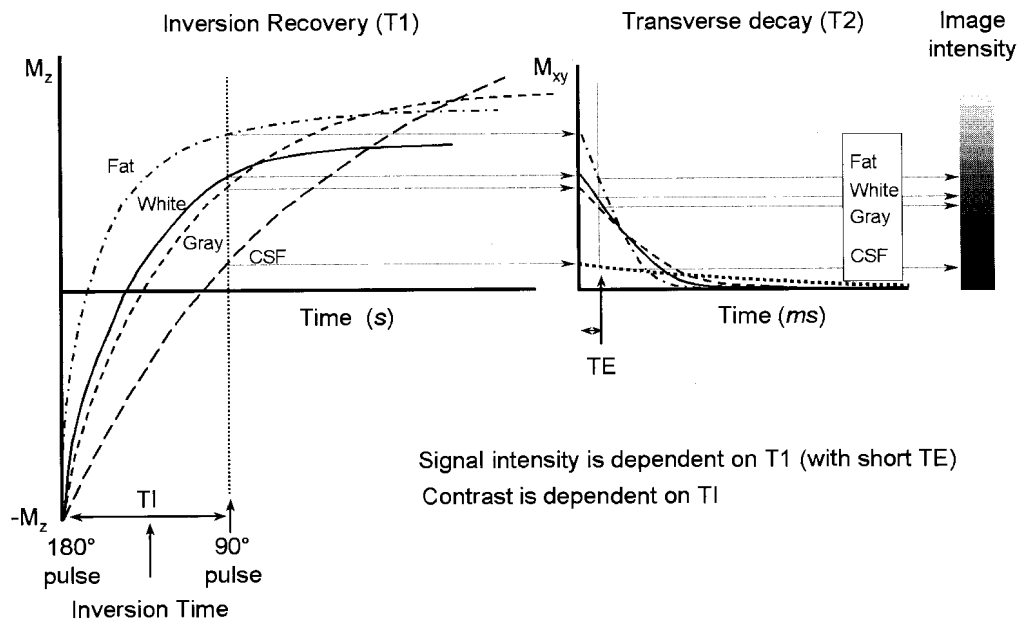


FIGURE 14-26. The inversion recovery longitudinal recovery diagram shows the $2 \times M_z$ range provided by the 180-degree excitation pulse. A 90-degree readout pulse at time TI and a 180-degree refocusing pulse at time TE/2 from the readout pulse form the echo at time TE. The time scale is not explicitly indicated on the x-axis, nor is the 180-degree refocusing pulse. A short TE is used to reduce T2 contrast characteristics.

tization *when a short TI is used*. The actual signals are encoded in one of two ways: (a) a *magnitude* (absolute value) signal flips negative M_z values to positive values, or (b) a *phase* signal preserves the full range of the longitudinal magnetization from $-M_z$ to $+M_z$, and the relative signal amplitude is displayed from dark to bright across the dynamic range of the image display. The first (magnitude) method is used most often, for reasons described later.

Short Tau Inversion Recovery

Short tau inversion recovery, or STIR, is a pulse sequence that uses a *very short TI* and magnitude signal processing (Fig. 14-27A). In this sequence, there are two interesting outcomes: (a) materials with short T1 have a *lower* signal intensity (the reverse of a standard T1-weighted image), and (b) all tissues pass through zero amplitude ($M_z = 0$), known as the *bounce point* or tissue null. Because the time required to reach the bounce point is known for a variety of tissue types, judicious TI selection can suppress a given tissue signal (e.g., fat).

Analysis of the equations describing longitudinal recovery with this excitation reveals that the signal *null* ($M_z = 0$) occurs at $TI = \ln(2) \times T1$, where $\ln$ is the natural log and $\ln(2) = 0.693$. For fat suppression, the null occurs at a time equal to $0.693 \times 260 \text{ msec} = 180 \text{ msec}$ (where $T1 \approx 260 \text{ msec}$ at 1.5 T). A typical STIR sequence uses a TI of 140 to 180 msec and a TR of 2,500 msec. Compared with a T1-weighted examination, STIR “fat suppression” reduces distracting fat signals (see Fig. 14-27B) and eliminates chemical shift artifacts (explained in Chapter 15).

Fluid Attenuated Inversion Recovery

The use of a longer TI reduces the signal levels of CSF and other tissues with long T1 relaxation constants, which can be overwhelming in the magnitude inversion recovery image. *Fluid attenuated inversion recovery*, the *FLAIR* sequence, reduces CSF signal and other water-bound anatomy in the MR image by using a TI selected at or near the bounce point of CSF. This permits a better evaluation of the anatomy that has a significant fraction of fluids with long T1 relaxation constants. Nulling of the CSF signal requires $TI = \ln(2) \times 3,500 \text{ msec}$, or about 2,400 msec. A TR of 7,000 msec or longer is typically employed to allow reasonable M_z recovery (see Fig. 14-27C).

Contrast Comparison of the Spin Echo Sequences

MR contrast schemes for clinical imaging use the spin echo or a variant of the spin echo pulse sequences for most examinations. Comparisons of the contrast characteristics are essential for accurate differential diagnoses and determination of the extent of pathology. This is illustrated with brain images obtained with T1, proton density, T2, and FLAIR contrast sequences (Fig. 14-28).

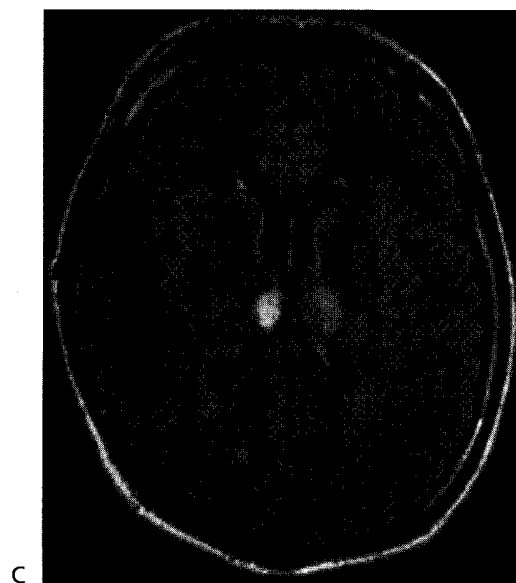
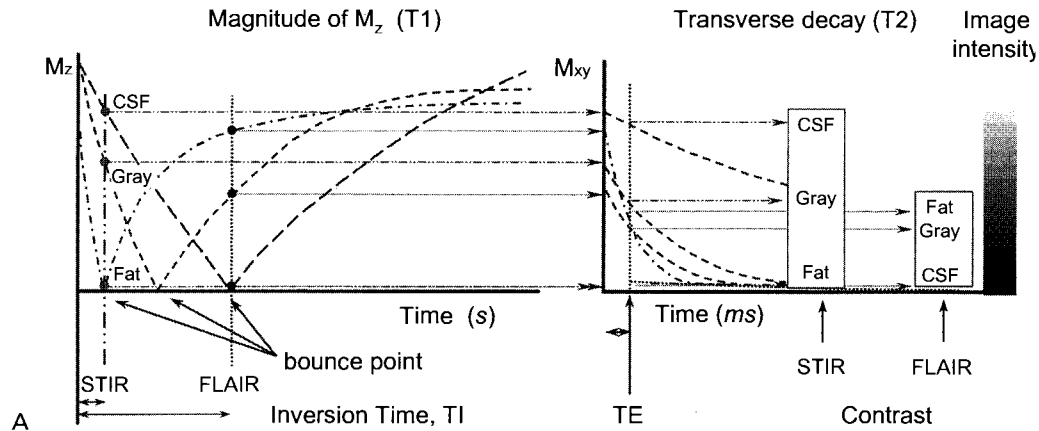


FIGURE 14-27. A: The short tau inversion recovery (STIR) and fluid attenuation inversion recovery (FLAIR) techniques use inversion recovery with magnitude longitudinal magnetization signals. This allows tissues to be nulled at the "bounce point" and is a method by which to apply selective tissue suppression based on inversion time (TI) and T1 values. Note the reduced fat signal for STIR and the reduced cerebrospinal fluid (CSF) signal for FLAIR. **B:** A STIR sequence (TR = 5,520 msec, TI = 150 msec, TE = 27.8 msec) and comparison T1 sequence (TR = 750 msec, TE = 13.1 msec) shows the elimination of fat signal. **C:** A FLAIR sequence (TR = 10,000 msec, TI = 2,400 msec, TE = 150 msec) shows the suppression of CSF in the image, caused by selection of a TI with which CSF is de-emphasized.



FIGURE 14-28. Magnetic resonance sequences produce variations in contrast for delineation of anatomy and pathology, including T1, proton density (PD), T2, and fluid attenuation inversion recovery (FLAIR) sequences. The images illustrate the characteristics of T1 with high fat signal; PD with high signal-to-noise ratio (SNR) yet low contrast discrimination; T2 with high signal and contrast for long-T2 tissues; and FLAIR, with fluid-containing tissues clearly delineated as a low signal. All images assist in the differential diagnosis of disease processes.

14.6 GRADIENT RECALLED ECHO

The *gradient recalled echo* (GRE) technique uses a magnetic field gradient to induce the formation of an echo, instead of the 180-degree pulses discussed earlier. A gradient magnetic field changes the local magnetic field to slightly higher and slightly lower strength and therefore causes the proton frequencies under its influence to precess at slightly higher and lower frequencies along the direction of the applied gradient. For an FID signal generated under a linear gradient, the transverse magnetization dephases rapidly as the gradient is continually applied. If the gradient polarity is instantaneously reversed after a predetermined time, the spins will rephase and produce a *gradient echo*; its peak amplitude occurs when the opposite gradient (of equal strength) has been applied for the same time as the initial gradient. Continually applying this latter (opposite) gradient allows the echo to decay, during which time the signal can be acquired (Fig. 14-29).

The GRE is *not* a true spin echo technique but a *purposeful dephasing and rephasing of the FID*. Magnetic field (B_0) inhomogeneities and tissue susceptibilities (magnetic field inhomogeneities caused by paramagnetic or diamagnetic tissues or contrast agents) are emphasized in GRE, because the dephasing and rephasing occur in the *same* direction as the main magnetic field and *do not* cancel the inhomogeneity effects, as the 180-degree refocusing RF pulse does. Figures 14-17 and 14-29 (the rotating frame diagram) may be compared to illustrate the direction of the M_{xy} vector of the FID and the echo. With spin echo techniques, the FID and echo vectors form in opposite directions, whereas with GRE they form in the same direction. Unlike the image produced by spin echo techniques, the GRE sequence has a significant sensitivity to field nonuniformity and magnetic susceptibility agents. The decay of transverse magnetization is therefore a strong function of $T2^*$, which is much shorter than the “true” T2 seen in spin echo sequences. The “echo” time is controlled either inserting a time delay between the negative and positive gradients or by reducing the amplitude of the

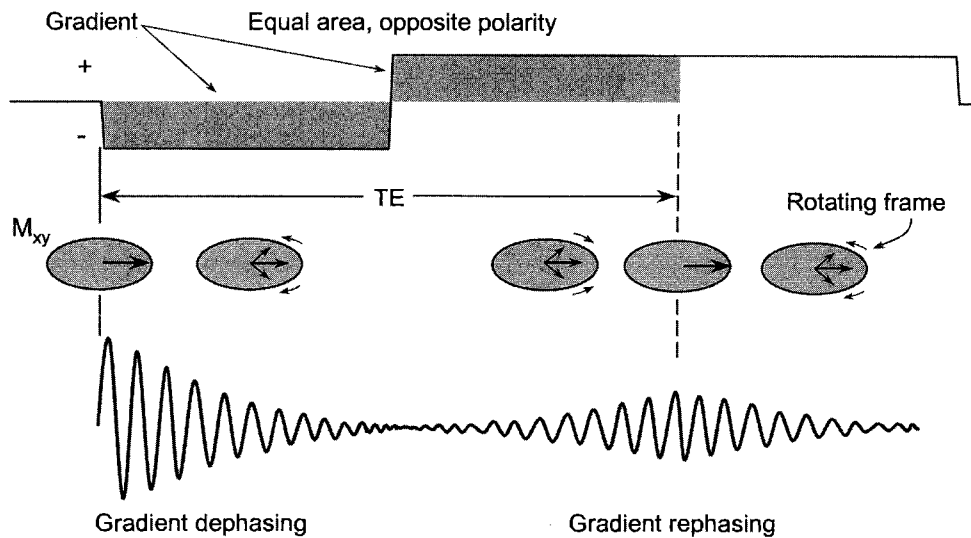


FIGURE 14-29. Gradient imaging techniques use a magnetic field gradient instead of a 180-degree radiofrequency pulse to induce the formation of an "echo." Dephasing the transverse magnetization spins with an applied gradient of one polarity and rephasing the spins with the gradient reversed in polarity produces a "gradient recalled echo." Note that the rotating frame depicts the magnetic moment vector of the echo in the *same direction* as the free induction decay (FID) relative to the main magnetic field, and therefore extrinsic inhomogeneities are not cancelled.

reversed (usually positive) gradient, thereby extending the time for the rephasing process to occur.

A major variable determining tissue contrast in GRE sequences is the flip angle. Depending on the desired contrast, flip angles of a few degrees to more than 90 degrees are applied. With a short TR, smaller flip angles are used because there is less time to excite the spin system, and *more* transverse magnetization is actually

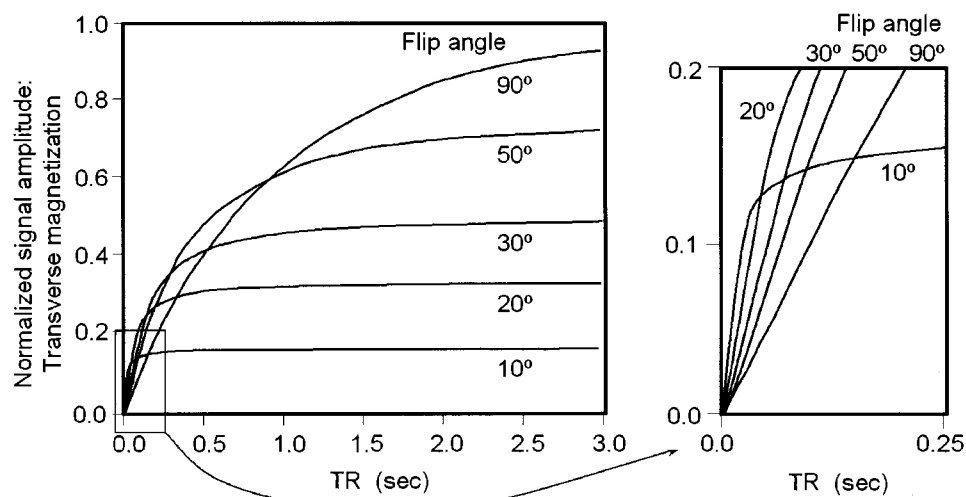


FIGURE 14-30. Transverse magnetization as a function of TR and the flip angle. For small flip angles and short TR, the transverse magnetization is actually higher than for larger flip angles.

generated with smaller flip angles than with a larger flip angle. This occurs because magnetization quickly builds up in the tissue sample. Figure 14-30 shows a plot of the signal amplitude of the transverse magnetization versus TR as a function of flip angle. Notice the lower signal amplitude for the larger flip angles with TR less than 200 msec. Tissue contrast in GRE pulse sequences depends on TR, TE, and the flip angle, as discussed in the following sections.

Gradient Recalled Echo Sequence with Long TR (>200 msec)

For the GRE sequence with “long” TR (>200 msec) and *flip angles greater than 45 degrees*, contrast behavior is similar to that of spin echo sequences. The major difference is the generation of signals based on T2* contrast rather than on T2 contrast, because the gradient echo does not cancel B₀ inhomogeneities, as a spin echo technique does. In terms of clinical interpretation, the mechanisms of T2* contrast are different from those of T2, particularly for MRI contrast agents. A long TE tends to show the *differences* between T2* and T2 rather than the improvement of T2 contrast, as would be expected in a spin echo sequence. T1 weighting is achieved with a short TE (less than 15 msec).

For flip angles less than 30 degrees, the small transverse magnetization amplitude reduces the T1 differences among the tissues. Proton density differences are the major contrast attributes for short TE, whereas longer TE provides T2* weighting. In most situations, GRE imaging is not useful with the relatively long TR, except when contrast produced by magnetic susceptibility differences is desired.

Steady-State Precession with Short TR (<50 msec)

Steady-state precession refers to equilibration of the *longitudinal and transverse magnetization* from pulse to pulse in an image acquisition sequence. In all standard imaging techniques (e.g., spin echo, inversion recovery, generic GRE sequences), the repetition period TR of RF pulses is too short to allow recovery of longitudinal magnetization equilibrium, and thus partial saturation occurs. When this steady-state partial saturation occurs, the *same* longitudinal magnetization is present for each subsequent pulse. For *very* short TR, *less than* the T2* decay constant, persistent *transverse magnetization* also occurs. During each pulse sequence repetition a portion of the transverse magnetization is converted to longitudinal magnetization *and* a portion of the longitudinal magnetization is converted to transverse magnetization. Under these conditions, *steady-state longitudinal and transverse magnetization components* coexist at all times in a dynamic equilibrium. Acronyms for this situation include GRASS (gradient recalled acquisition in the steady state), FISP (fast imaging with steady-state precession), FAST (Fourier acquired steady state), and other acronyms given by the MRI equipment manufacturers.

The user-defined factors that determine the amount of transverse magnetization created by each RF pulse and the image contrast in steady-state imaging include TR, TE, and flip angle. Steady-state imaging is practical only with *short* and

very short TR. In these two regimes, flip angle has the major impact on the contrast “weighting” of the resultant images.

Small Flip Angles. With small flip angles from 5 to 30 degrees, the amplitude of longitudinal magnetization is very small, resulting in a very small transverse magnetization. Therefore, T1 relaxation differences between tissues are small. For the same reasons, T2* decay differences are small, particularly with a short TE. Spin density weighting is the single most important factor for short TR, small flip angle techniques.

Moderate Flip Angles. With moderate flip angles of 30 to 60 degrees, a larger longitudinal magnetization produces a larger transverse magnetization, and tissues with a long T1 have more signal saturation. This produces some T1 weighting. For the same reason, tissues with long T2* generate a higher signal amplitude. Because T1 and T2 for most tissues are correlated (i.e., long T1 implies long T2), the contrast depends on the difference in T2/T1 ratio between the tissues. For example, parenchyma tissue contrast is reduced (e.g., white versus gray matter), because the T2/T1 ratios are approximately equal. On the other hand, the T2/T1 ratio for CSF is relatively high, which generates a large signal difference compared with the parenchyma and produces a larger signal.

Large Flip Angles. For flip angles in the range of 75 to 90 degrees, almost complete saturation of the longitudinal magnetization occurs, reducing the T1 tissue contrast. Conversely, a large transverse steady-state magnetization produces greater T2* weighting and reduces spin density differences. The contrast for short TR, large flip angle techniques depends on T2* and T1 weighting.

The echo time, TE, also influences the contrast. The descriptions of flip angle influence assume a very short TE (e.g., less than 5 msec). As TE increases, a greater difference in T2* decay constants occurs and contributes to the T2* contrast in the image.

Steady-State Gradient Recalled Echo Contrast Weighting

Table 14-6 indicates typical parameter values for contrast desired in steady state and GRE acquisitions. The steady-state sequences produce T2* and spin density contrast. A GRASS/FISP sequence using TR = 35 msec, TE = 3 msec, and flip angle = 20 degrees shows unremarkable contrast, but blood flow shows up as a

TABLE 14-6. GRADIENT RECALLED ECHO WEIGHTING (STEADY-STATE)

Parameter	T1	T2/T1	T2	T2*	Spin Density
Flip angle (degrees)	45–90	30–50	5–15	5–15	5–30
TR (msec)	200–400	10–50	200–400	100–300	100–300
TE (msec)	3–15	3–15	30–50	10–20	5–15



FIGURE 14-31. A steady-state gradient recalled echo (GRASS) sequence (TR = 24 msec, TE = 4.7 msec, flip angle = 50 degrees) of two slices out of a volume acquisition is shown. Contrast is unremarkable for white and gray matter because of a T2/T1 weighting dependence. Blood flow appears as a relatively bright signal. MR angiography depends on pulse sequences such as these to reduce the contrast of the anatomy relative to the vasculature.

bright signal (Fig. 14-31). This technique enhances vascular contrast, from which MR angiography sequences can be reconstructed (see Chapter 15).

“Spoiled” Gradient Echo Techniques

With very short TR steady-state acquisitions, T1 weighting *cannot* be achieved to any great extent, owing to either a small difference in longitudinal magnetization with small flip angles, or dominance of the T2* effects for larger flip angles produced by persistent transverse magnetization. The T2* influence can be reduced by using a long TR (usually not an option), or by “spoiling” (destroying) the steady-state transverse magnetization by adding a phase shift to successive RF pulses during the acquisition. (Both the transmitter and receiver must be phase locked for this approach to work.) By shifting the residual transverse components out of phase, the buildup of the transverse steady-state signal does not occur, effectively eliminating the T2* dependency of the signal. Although small T2* contributions are introduced by the gradient reversal, mostly *T1-weighted contrast* is obtained. Short TR, short TE, moderate to large flip angle, and *spoiled* transverse magnetization produces the greatest T1 contrast. Currently, spoiled transverse magnetization gradient recalled echo (SPGR) is used in three-dimensional volume acquisitions because of the extremely short acquisition time allowed by the short TR of the GRE sequence and the good contrast rendition of the anatomy provided by T1 weighting (Fig. 14-32).

MR contrast agents (e.g., gadolinium) produce greater contrast with T1-weighted spoiled GRE techniques than with a comparable T1-weighted spin echo sequence because of the greater sensitivity to magnetic susceptibility of the contrast agent. The downside of spoiled GRE techniques is the increased sensitivity to other artifacts such as chemical shift and magnetic field inhomogeneities, as well as a lower SNR.

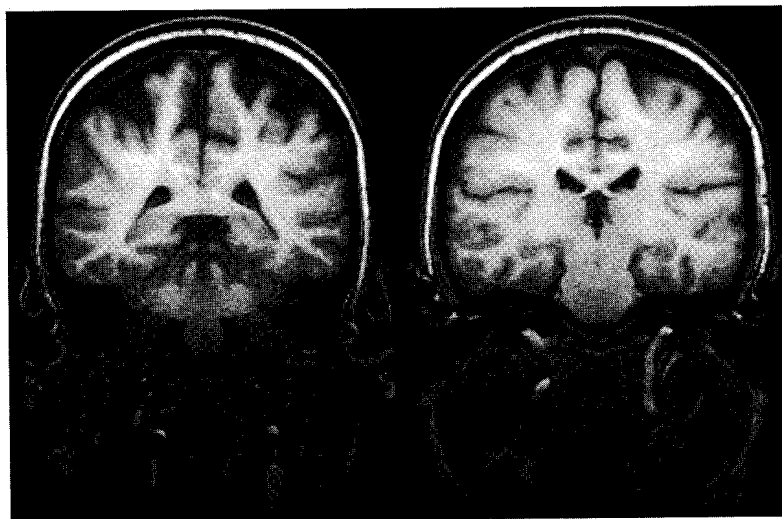


FIGURE 14-32. A spoiled transverse magnetization gradient recalled echo (SPGR) sequence (TR = 8 msec, TE = 1.9 msec, flip angle = 20 degrees) of two slices out of a volume acquisition is shown. The ability to achieve T1 contrast weighting is extremely useful for rapid three-dimensional volume imaging. Bright blood (lower portion of each image) and magnetic susceptibility artifacts are characteristic of this sequence.

14.7 SIGNAL FROM FLOW

Moving fluid (vascular and CSF) has an appearance in MR images that is complicated by many factors, including flow velocity, vessel orientation, laminar versus turbulent flow patterns, pulse sequences, and image acquisition modes. Flow-related mechanisms combine with image acquisition parameters to alter contrast. Signal due to flow covers the entire gray scale of MR signal intensities, from “black-blood” to “bright-blood” levels, and flow can be a source of artifacts. The signal from flow can also be exploited to produce MR angiographic images.

Low signal intensities (flow voids) are often a result of *high-velocity signal loss* (HVSL), in which protons in the flowing blood move out of the slice during echo reformation, causing a lower signal (see discussion of slice selection methods in Chapter 15). *Flow turbulence* can also cause flow voids, by causing a dephasing of spins in the blood with a resulting loss of the magnetic moment in the area of turbulence. With HVSL, the amount of signal loss depends on the velocity of the moving fluid. Pulse sequences to produce “black blood” in images can be very useful in cardiac and vascular imaging. A typical black-blood pulse sequence uses a “double inversion recovery” method, whereby a nonselective 180-degree RF pulse is initially applied, inverting all spins in the body, and is followed by a selective 180-degree RF pulse that restores the magnetization in the selected slice. During the inversion time, blood outside of the excited slice with inverted spins flows into the slice, producing no signal; therefore, the blood appears dark.

Flow-related enhancement is a process that causes increased signal intensity due to flowing spins; it occurs during imaging of a volume of tissues (see Chapter 15).

Even-echo rephasing is a phenomenon that causes flow to exhibit increased signal on even echoes in a multiple-echo image acquisition. Flowing spins that experience two subsequent 180-degree pulses (even echoes) generate higher signal intensity due to a constructive rephasing of spins during echo formation. This effect is prominent in slow laminar flow (e.g., veins show up as bright structures on even-echo images).

Flow enhancement in gradient echo images is pronounced for both venous and arterial structures, as well as CSF. The high intensity is caused by the wash-in (between subsequent RF excitations) of fully unsaturated spins (with full magnetization) into a volume of partially saturated spins caused by the short TR used with gradient imaging. During the next excitation, the signal amplitude resulting from the moving unsaturated spins is about 10 times greater than that of the nonmoving saturated spins. With gradient echo techniques, the degree of enhancement depends on the velocity of the blood, the slice or volume thickness, and the TR. As blood velocity increases, unsaturated blood exhibits the greatest signal. Similarly, a thinner slice or decreased repetition time results in higher flow enhancement. In arterial imaging of high-velocity flow, it is possible to have bright blood throughout the imaging volume of a three-dimensional acquisition if unsaturated blood can penetrate into the volume without experiencing an RF pulse.

14.8 PERFUSION AND DIFFUSION CONTRAST

Perfusion of tissues via the capillary bed permit the delivery of oxygen and nutrients to the cells and removal of waste (e.g., CO₂) from the cells. Conventional perfusion measurements are based on the uptake and wash-out of radioactive tracers or other exogenous tracers that can be quantified from image sequences using well-characterized imaging equipment and calibration procedures. For MR perfusion images, exogenous and endogenous tracer methods are used. Freely diffusible tracers using nuclei such as ²H (deuterium), ³He, ¹⁷O, and ¹⁹F are employed in experimental procedures to produce differential contrast in the tissues. More clinically relevant are intravascular blood-pool agents such as gadolinium–diethylenetriaminepentaacetic acid (Gd-DTPA), which modify the relaxation of protons in the blood in addition to producing a shorter T2*. *Endogenous* tracer methods do not use externally added agents, but instead depend on the ability to generate contrast from specific excitation or diffusion mechanisms. For example, labeling of inflowing spins (“black-blood” perfusion) uses protons in the blood as the contrast agent. Tagged (labeled) spins outside of the imaging volume perfuse into the tissues, resulting in a drop in signal intensity, a time course of events that can be monitored by quantitative measurements.

Blood oxygen level–dependent (BOLD) and “functional MR” imaging rely on the differential contrast generated by blood metabolism in active areas of the brain. In flow of blood and conversion of oxyhemoglobin to deoxyhemoglobin (a paramagnetic agent) increases the magnetic susceptibility in the localized area and induces signal loss by reducing T2*. This fact allows areas of high metabolic activity to produce a correlated signal and is the basis of functional MRI (fMRI). A BOLD sequence produces an image of the head before the application of the stimulus. The patient is subjected to the stimulus and a second BOLD image is

acquired. Because the BOLD sequence produces images that are highly dependent on blood oxygen levels, areas of high metabolic activity will be enhanced when the *prestimulus image* is subtracted, pixel by pixel, from the poststimulus image. These fMRI experiments determine which sites in the brain are used for processing data. Stimuli in fMRI experiments can be physical (finger movement), sensory (light flashes or sounds), or cognitive (repetition of “good” or “bad” word sequences). To improve the SNR in the fMRI images, a stimulus is typically applied in a repetitive, periodic sequence, and BOLD images are acquired continuously. Areas in the brain that demonstrate time-dependent activity and correlate with the time-dependent application of the stimulus are coded using a color scale and are overlaid onto a gray-scale image of the brain for anatomic reference.

Diffusion depends on the random motion of water molecules in tissues. Interaction of the local cellular structure with the diffusing water molecules produces anisotropic, directionally dependent diffusion (e.g., in the white matter of brain tissues). Diffusion-weighted sequences use a strong MR gradient (see Chapter 15) applied to the tissues to produce signal differences based on the mobility and directionality of water diffusion. Tissues with more water mobility have a greater signal

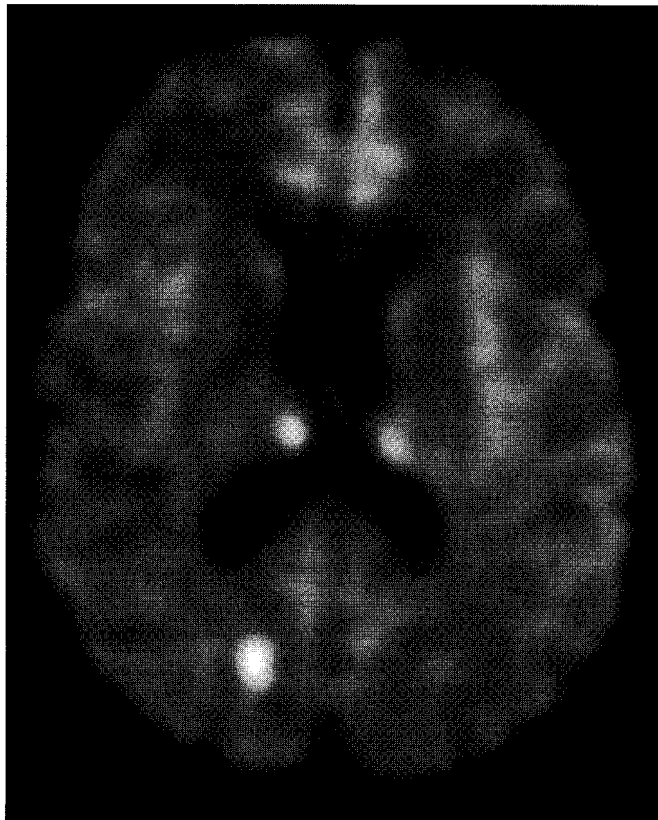


FIGURE 14-33. Diffusion-weighted image contrast illustrates gray scale-encoded diffusion coefficients of the brain. (This image corresponds to the images in Figs. 14-21 through 14-23 and 14-27C).

loss than those of lesser water mobility. The *in vivo* structural integrity of certain tissues (healthy, diseased, or injured) can be measured by the use of diffusion-weighted imaging (DWI), in which water diffusion characteristics are determined through the generation of apparent diffusion coefficient (ADC) maps (Fig. 14-33). ADC maps of the spine and the spinal cord have shown promise in predicting and evaluating pathophysiology before it is visible on conventional T1- or T2-weighted images. DWI is also a sensitive indicator for early detection of ischemic injury. For instance, areas of acute stroke show a drastic reduction in the diffusion coefficient compared with nonischemic tissues. Diagnosis of multiple sclerosis is a potential application, based on the measurements of the diffusion coefficients in three-dimensional space.

Various acquisition techniques are used to generate diffusion-weighted contrast. Standard spin-echo and echoplanar image (EPI, see Chapter 15) pulse sequences with diffusion gradients are common. Obstacles to diffusion weighting are the extreme sensitivity to motion of the head and brain, which is chiefly caused by the large pulsed gradients required for the diffusion preparation, and eddy currents, which reduce the effectiveness of the gradient fields. Several strategies have been devised to overcome the motion sensitivity problem, including common electrocardiographic gating, motion compensation methods, and eddy current compensation.

14.9 MAGNETIZATION TRANSFER CONTRAST

Magnetization transfer contrast is the result of selective observation of the interaction between protons in free water molecules and protons contained in the macromolecules of a protein. Magnetization exchange occurs between the two proton groups because of coupling or chemical exchange. Because the protons exist in slightly different magnetic environments, the selective saturation of the protons in the macromolecule can be excited *separately* from the bulk water by using narrow-band RF pulses (because the Larmor frequencies are different). A *transfer of the magnetization* from the protons in the macromolecule partially saturates the protons in bulk water, even though these protons have not experienced an RF excitation pulse. Reduced signal from the adjacent free water protons by the saturation “label” affects *only those protons having a chemical exchange with the macromolecules* and improves image contrast in many situations (Fig. 14-34). This technique is used for anatomic MR imaging of the heart, the eye, multiple sclerosis, knee cartilage, and general MR angiography. Tissue characterization is also possible, because the image contrast in part is caused by the surface chemistry of the macromolecule and the tissue-specific factors that affect the magnetization transfer characteristics.

MR pulse sequences allow specific contrast differences between the tissues to be generated by manipulating the magnetization of the tissues in well-defined ways. The next chapter (Chapter 15) discusses the methods of signal acquisition and image reconstruction, as well as image characteristics, artifacts, quality control, and magnetic field safety and bioeffects.

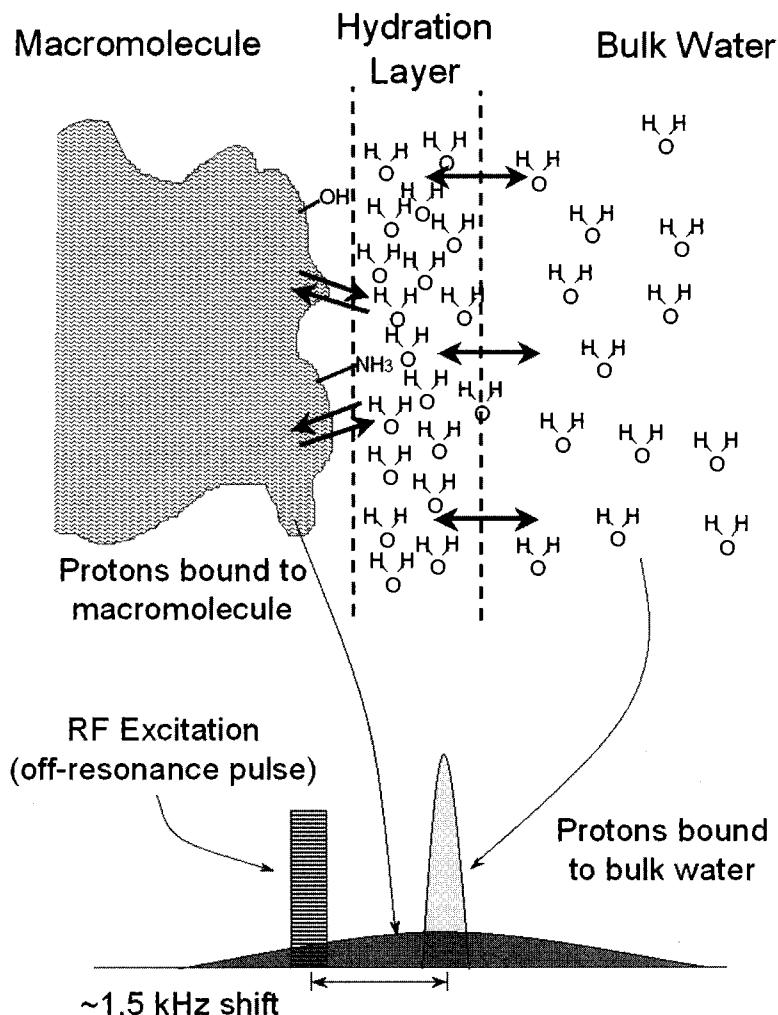


FIGURE 14-34. Magnetization transfer contrast is implemented with an off-resonance radiofrequency pulse of about 1,500 Hz from the Larmor frequency. Excitation and dephasing of hydrogen atoms on the surface of macromolecules is transferred via the hydration layer to adjacent "free water" hydrogen atoms. A partial saturation of the tissues reduces the signals that would otherwise compete with signals from blood flow.

SUGGESTED READING

- Axel L, et al. *Glossary of MR terms*, 3rd ed. Reston, VA: American College of Radiology, 1995.
- Brown MA, Smelka RC. *MR: basic principles and applications*. New York: John Wiley & Sons, 1995.
- Hendrick RE. The AAPM/RSNA physics tutorial for residents. Basic physics of MR imaging: an introduction. *Radiographics* 1994;14:829–846.
- NessAiver M. *All you really need to know about MRI physics*. Baltimore: Simply Physics, 1997.
- Plewes DB. The AAPM/RSNA physics tutorial for residents. Contrast mechanisms in spin-echo MR imaging. *Radiographics* 1994;14:1389–1404.

- Price RR. The AAPM/RSNA physics tutorial for residents. Contrast mechanisms in gradient-echo imaging and an introduction to fast imaging. *Radiographics* 1995;15:165–178.
- Smith HJ, Ranallo FN. *A non-mathematical approach to basic MRI*. Madison, WI: Medical Physics Publishing, 1989.
- Smith RC, Lange RC. *Understanding magnetic resonance imaging*. Boca Raton, FL: CRC Press, 1998.
- Wherli FW. *Fast-scan magnetic resonance: principles and applications*. New York: Raven Press, 1991.

MAGNETIC RESONANCE IMAGING (MRI)

The importance of magnetic resonance imaging (MRI) in clinical imaging has exceeded even the most optimistic hopes of researchers from the 1980s. An ability to manipulate and adjust tissue contrast with increasingly complex pulse sequences is at the crux of this success, as described in the previous chapter. However, the ability to accurately determine the position from the nuclear magnetic resonance (NMR) signal and thus create an image is the basis of clinical MRI. This chapter describes how the MR signals are localized in three dimensions, and the relationship between frequency and spatial position is examined. Characteristics of the MR image, including factors that affect the signal-to-noise ratio (SNR), blood flow effects, and artifacts are described. Concluding the chapter is an overview of MR equipment, quality control procedures, and biologic safety issues with respect to magnetic fields and radiofrequency (RF) energy.

15.1 LOCALIZATION OF THE MR SIGNAL

The protons in a material, with the use of an external uniform magnetic field and RF energy of specific frequency, are excited and subsequently produce signals with amplitudes dependent on relaxation characteristics and spin density, as previously discussed (see Chapter 14). Spatial localization, fundamental to MR imaging, requires the imposition of magnetic field nonuniformities—magnetic gradients superimposed upon the homogeneous and much stronger main magnetic field, which are used to distinguish the positions of the signal in a three-dimensional object (the patient). Conventional MRI involves RF excitations combined with magnetic field gradients to localize the signal from individual volume elements (voxels) in the patient.

Magnetic Field Gradients

Magnetic fields with predictable directionality and strength are produced in a coil wire energized with a direct electric current of specific polarity and amplitude. Magnetic field gradients are obtained by superimposing the magnetic fields of one or more coils with a precisely defined geometry (Fig. 15-1). With appropriate design, the gradient coils create a magnetic field that linearly varies in strength versus distance over a pre-

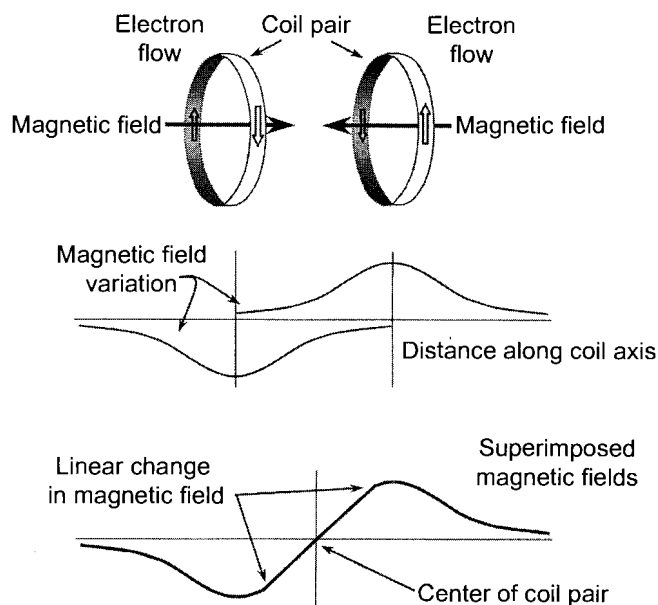


FIGURE 15-1. Individual conducting wire coils that are separately energized with currents of opposite directions produce magnetic fields of opposite polarity. Magnetic field strength reduces with distance from the center of each coil. When combined, the magnetic field variations form a linear change between the coils, producing a linear magnetic field gradient.

defined field of view (FOV). When superimposed upon a homogeneous magnetic field (e.g., the main magnet, B_0), positive gradient field adds to B_0 and negative gradient field reduces B_0 . Inside the magnet bore, three sets of gradients reside along the coordinate axes— x , y , and z —and produce a magnetic field variation determined by the magnitude of the applied current in each coil set (Fig. 15-2). When independently energized, the three coils (x , y , z) produce a linearly variable magnetic field in any direction, where the net gradient is equal to $\sqrt{G_x^2 + G_y^2 + G_z^2}$. Gradient polarity reversals (positive to negative and negative to positive changes in magnetic field strength) are achieved by reversing the current direction in the gradient coils.

Two properties of gradient systems are important: (a) The peak amplitude of the gradient field determines the “steepness” of the gradient field. Gradient magnetic field strength typically ranges from 1 to 50 millitesla per meter (mT/m) [0.1 to 5 gauss (G)/cm]. (b) The slew rate is the time required to achieve the peak magnetic field amplitude, where shorter time is better. Typical slew rates of gradient fields are from 5 mT/m/msec to 250 mT/m/msec. Limitations in the slew rate are caused by eddy currents, which are electric currents induced in nearby conductors that oppose the generation of the gradient field (e.g., both the RF coils and the patient produce eddy currents). Actively shielded gradient coils and compensation circuits reduce the problems introduced by eddy currents.

The gradient is a linear, position-dependent magnetic field applied across the FOV, and it causes protons to alter their precessional frequency corresponding to their position along the applied gradient in a known and predictable way. At the middle of the gradient exists the null, where no change in the net magnetic field or precessional frequency exists. A linear increase (or decrease) in precessional frequency occurs with the variation of the local magnetic field strength away from the null (Fig. 15-3 and Table 15-1). Location of protons along the gradient is determined by their frequency and phase; the gradient amplitude and the number of samples over the FOV determine the frequency bandwidth (BW) across each pixel.

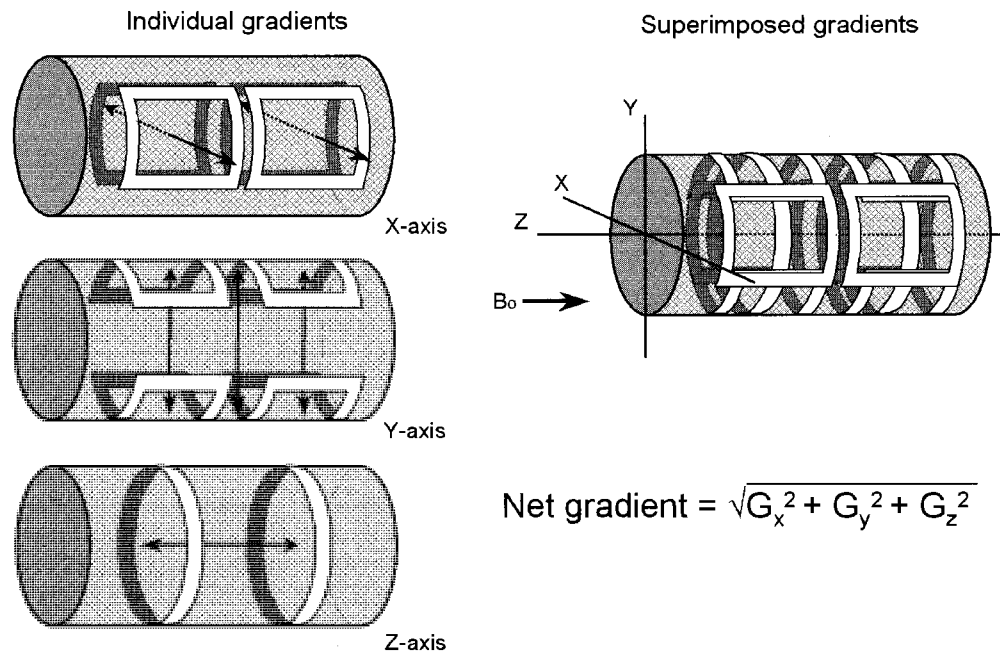


FIGURE 15-2. Within a large stationary magnetic field, field gradients are produced by three separate saddle coils placed within the central core of the magnet, along the x, y, or z direction. Magnetic field gradients of arbitrary direction are produced by the vector addition of the individual gradients turned on simultaneously. Any gradient direction is possible by superimposition of the three-axis gradient system.

The Larmor equation ($\omega = \gamma B$) allows the gradient amplitude (for protons) to be expressed in the units of Hz/cm. For instance, a 10-mT/m gradient can be expressed as $10 \text{ mT/m} \times 42.58 \text{ MHz/T} \times 1\text{T}/1,000 \text{ mT} = 0.4258 \text{ MHz/m}$, which is equivalent to 425.8 kHz/m or 4258 Hz/cm. This straightforward description of the gradient strength facilitates determining the frequency BW across the FOV, independent of the main magnet field strength.

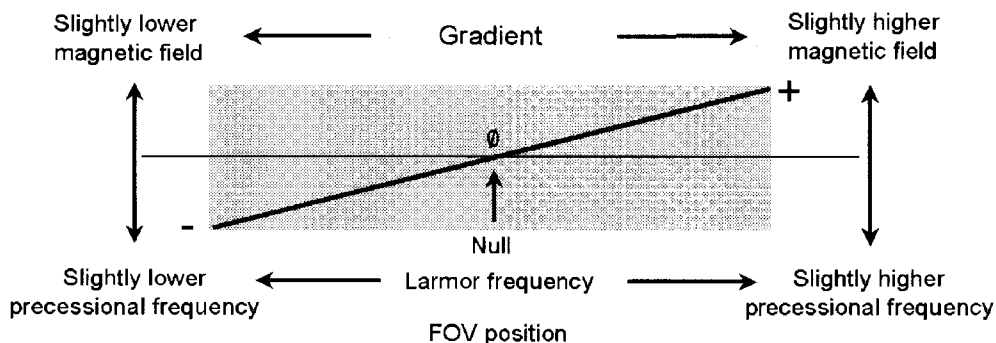


FIGURE 15-3. The gradient field creates a net positive and negative magnetic environment that adds to and subtracts from the main magnetic field. Associated with the local change in magnetic field is a local change in precessional frequencies, per the Larmor equation. The frequencies thus directly vary across the field in proportion to the applied gradient strength.

TABLE 15-1. PRECESSIONAL FREQUENCY VARIATION AT 1.5 T ALONG AN APPLIED GRADIENT

Gradient field strength	3 mT/m = 0.3 g/cm = 1,277.4 Hz/cm
Main magnetic field strength	1.5 T
Field of view (FOV)	15 cm
Linear gradient amplitude over FOV	0.45 mT; from -0.225 mT to +0.225 mT
Maximum magnetic field (frequency)	1.500225 T (63.8795805 MHz)
Unchanged magnetic field at null	1.500000 T (63.8700000 MHz)
Minimum magnetic field	1.499775 T (63.8604195 MHz)
Net frequency range across FOV ^a	0.019161 MHz = 19.2 kHz = 19,161 Hz
Frequency range across FOV (1,278 Hz/cm) ^b	1,277.4 Hz/cm $\times$ 15 cm = 19,161 Hz
Frequency bandwidth per pixel (256 samples)	19,161 Hz/256 = 74.85 Hz/pixel

^aCalculated using the absolute precessional frequency range: 63.8796–63.8604 MHz.

^bCalculated using the gradient strength expressed in Hz/cm.

The localization of protons in the three-dimensional volume requires the application of three distinct gradients during the pulse sequence: slice select, frequency encode, and phase encode gradients. These gradients are usually sequenced in a specific order, depending on the pulse sequences employed. Often, the three gradients overlap partially or completely during the scan to achieve a desired spin state, or to leave spins in their original phase state after the application of the gradient(s).

Slice Select Gradient

The RF antennas that produce the RF pulses do not have the ability to spatially direct the RF energy. Thus, the slice select gradient (SSG) determines the slice of tissue to be imaged in the body, in conjunction with the RF excitation pulse. For axial MR images, this gradient is applied along the long (cranial-caudal) axis of the body. Proton precessional frequencies vary according to their distance from the null of the SSG. A selective frequency (narrow band) RF pulse is applied to the whole volume, but only those spins along the gradient that have a precessional frequency equal to the frequency of the RF will absorb energy due to the resonance phenomenon (Fig. 15-4).

Slice thickness is determined by two parameters: (a) the bandwidth (BW) of the RF pulse, and (b) the gradient strength across the FOV. For a given gradient field

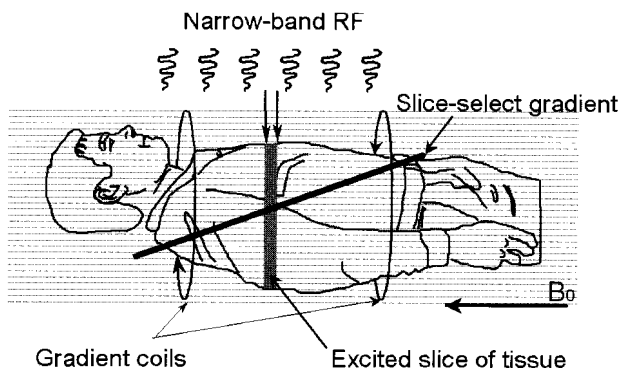


FIGURE 15-4. The slice select gradient (SSG) disperses the precessional frequencies of the protons in a known way along the gradient. A narrow-band radiofrequency (RF) pulse excites only a selected volume (slice) of tissues, determined by RF bandwidth and SSG strength.

strength, an applied RF pulse with a narrow BW excites the protons over a narrow slice of tissue, and a broad BW excites a thicker slice (Fig. 15-5A). For a fixed RF BW, the SEG field strength (slope) determines the slice thickness. An increase in the gradient produces a larger range of frequencies across the FOV and results in a decrease in the slice thickness (Fig. 15-5B).

The RF pulse used to excite a rectangular slice of protons requires the synthesis of a specialized waveform called a “sinc” pulse. The pulse contains a main lobe centered at “0” time and oscillations (negative and positive) that decrease in amplitude before and after the peak amplitude. To achieve a “perfect” slice profile (rectangular frequency BW), an infinite time is required before and after the pulse, an unachievable goal. Short pulse duration requires truncation of the sinc pulse, which produces less than ideal slice profiles. A longer duration RF pulse produces a better approximation to a desirable rectangular profile (Fig. 15-6). This is analogous to the concept of slice sensitivity profile in computed tomography (CT).

The sinc pulse width determines the output frequency BW. A narrow sinc pulse width and high-frequency oscillations produce a wide BW and a corresponding broad excitation distribution. Conversely, a broad, slowly varying sinc pulse produces a narrow BW and corresponding thin excitation distribution (Fig. 15-7).

A combination of a narrow BW and a low gradient strength or a wide BW and a high gradient strength can result in the same overall slice thickness. There are, however, considerations regarding the SNR of the *acquired* data as

$$\text{SNR} \propto \frac{1}{\sqrt{\text{BW}}}.$$

A narrow BW results in increased SNR. By decreasing the BW by a factor of four, the SNR is increased by a factor of two, as long as the same slice thickness is acquired by decreasing the gradient strength. Narrow BW is not always the appropriate choice, however, because chemical shift artifacts and other undesirable image

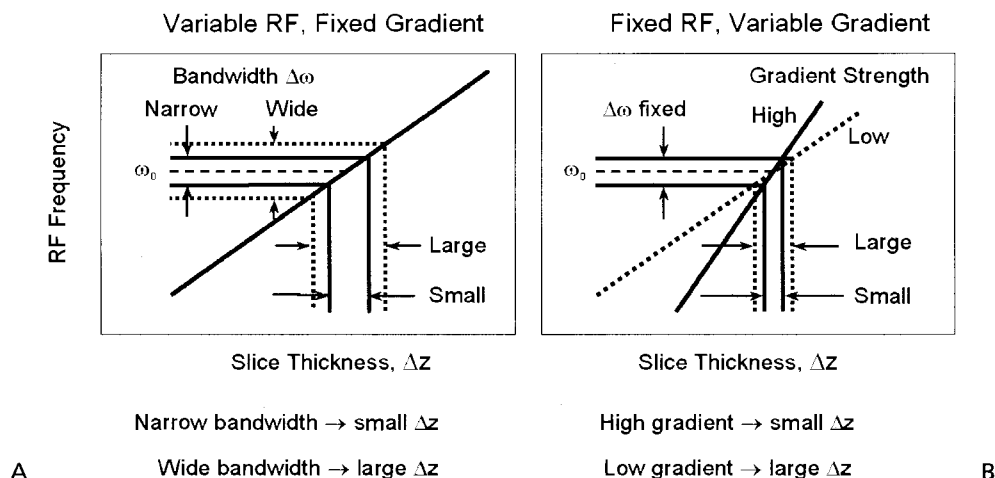


FIGURE 15-5. Slice thickness is dependent on RF bandwidth and gradient strength. **A:** For a fixed gradient strength, the RF bandwidth determines the slice thickness. **B:** For a fixed RF bandwidth, gradient strength determines the slice thickness.

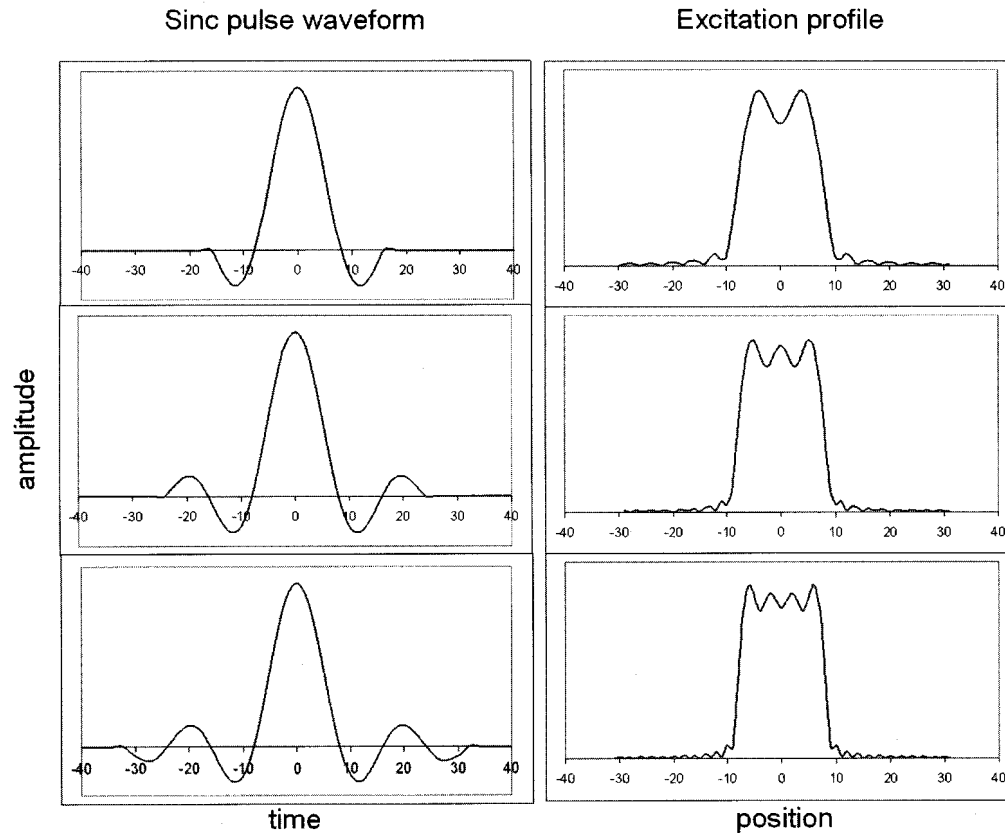


FIGURE 15-6. An approximate rectangular slice profile of precessing protons is achieved with the use of a RF waveform known as a “sinc” pulse, which evolves with oscillatory tails of increasing amplitude to the main lobe at time = 0, and decays in a symmetric manner with time. Truncation to limit the RF pulse duration produces nonrectangular slice profiles. As the RF pulse duration increases, the slice profile (inclusion of more lobes) is increased. The graphs have relative units for time, distance, and amplitude.

characteristics may result. Consequently, trade-offs in image quality must be considered when determining the optimal RF bandwidth and gradient field strength combinations for the SSG.

When applied, gradients induce spin dephasing across the patient imaging volume. To reestablish the original phase of all stationary protons after the slice-select excitation, a gradient of opposite polarity equal to one-half the area of the original gradient is applied (Fig. 15-8). For 180-degree RF excitations, the rephasing gradient is not necessary, as all spins maintain their phase relationships after the 180-degree pulse.

In summary, the slice select gradient applied during the RF pulse results in proton excitation in a single plane and thus localizes the signal in the dimension orthogonal to the gradient. It is the first of three gradients applied to the sample volume.

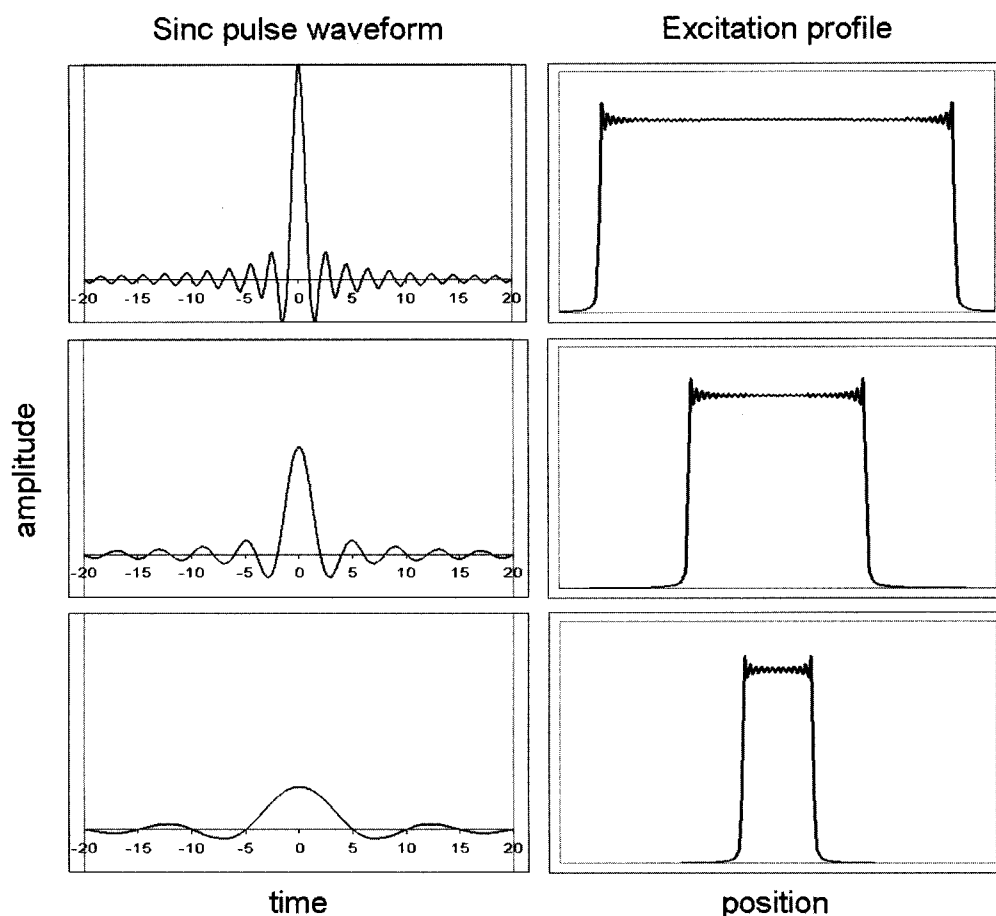


FIGURE 15-7. The width of the sinc pulse lobes determines the RF bandwidth. Narrow lobes produce wide bandwidth; wide lobes produce narrow bandwidth. In these examples, the area under the main sinc pulse lobe of each is approximately equal, resulting in equal amplitude of the corresponding spatial excitation characteristics. The graphs have relative units for time, distance and amplitude.

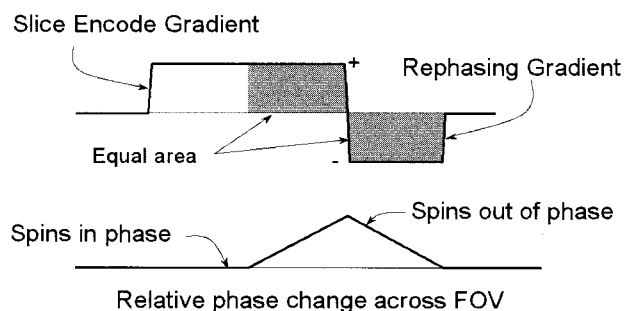


FIGURE 15-8. Gradients produce a range of precessional frequencies within the volume of tissue experiencing the magnetic field variations, which leaves the spins dephased after the gradient is removed. A rephasing gradient of opposite polarity serves to rewind the spins back to a coherent phase relationship when its area is equal to one-half of the total area of the original gradient.

Frequency Encode Gradient

The frequency encode gradient (FEG), also known as the readout gradient, is applied in a direction perpendicular to the SSG. For an axial image acquisition, the FEG is applied along the x-axis throughout the formation and the decay of the signals arising from the spins excited by the slice encode gradient. Spins constituting the signals are frequency encoded depending on their position along the FEG. During the time the gradient is turned on, the protons precess with a frequency determined by their position from the null. Higher precessional frequencies occur at the positive pole, and lower frequencies occur at the negative pole of the FEG. Demodulation (removal of the Larmor precessional frequency) of the composite signal produces a net frequency variation that is symmetrically distributed from 0 frequency at the null, to $+f_{\max}$ and $-f_{\max}$ at the edges of the FOV (Fig. 15-9). The composite signal is amplified, digitized, and decoded by the Fourier transform, a mathematical algorithm that converts frequency signals into a line of data corresponding to spin amplitude versus position (Fig. 15-10).

A spatial "projection" of the object is created by summing the signal amplitude along a column of the tissue; the width of the column is defined by the sampling aperture (pixel), and the thickness is governed by the slice select gradient strength

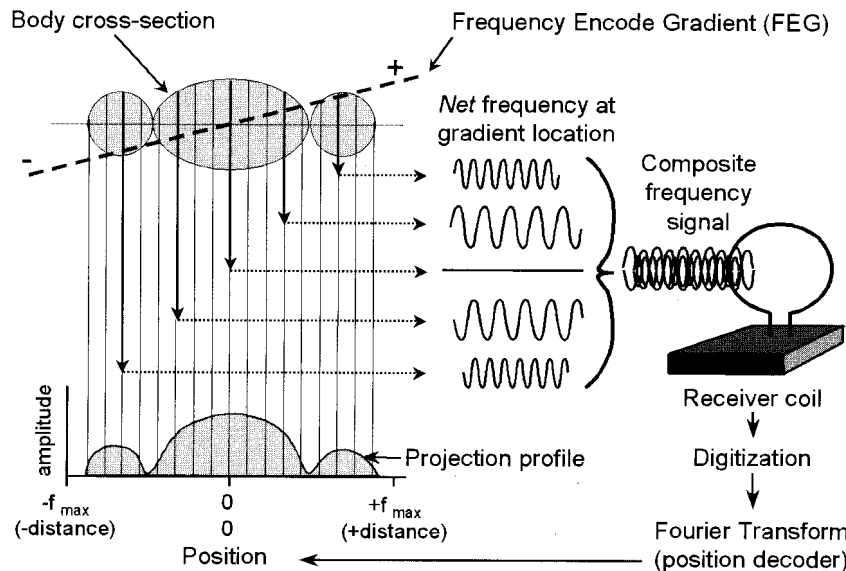


FIGURE 15-9. The frequency encode gradient (FEG) is applied in an orthogonal direction to the SSG, and confers a spatially dependent variation in the precessional frequencies of the protons. Acting only on those spins from the slice encode excitation, the composite signal is acquired, digitized, demodulated (Larmor frequency removed), and Fourier transformed into frequency and amplitude information. A one-dimensional array represents a projection of the slice of tissue (amplitude and position) at a specific angle. (Demodulation into net frequencies and amplitudes occurs after detection by the receiver coil; this is shown in the figure for clarity only.)

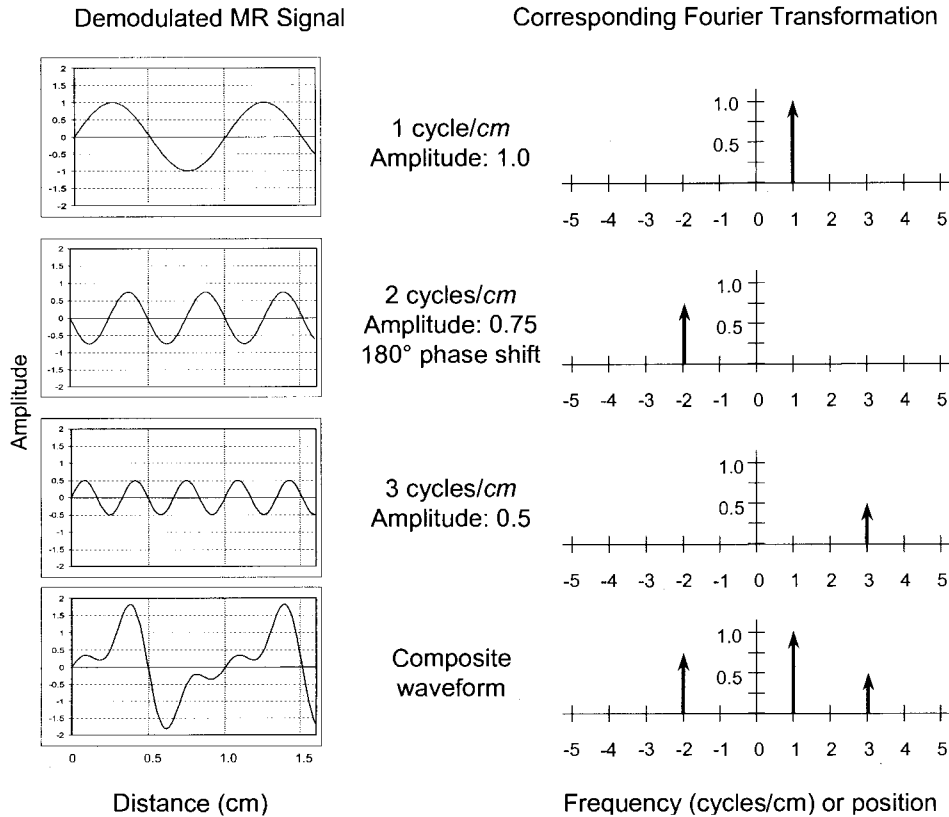


FIGURE 15-10. Spatial frequency signals (cycle/cm) and their Fourier transforms (spatial position) are shown for three simple sinusoidal waveforms with a specific amplitude and phase. The Fourier transform decodes the frequency, phase and amplitude variations in the spatial frequency domain into a corresponding position and amplitude in the spatial domain. A 180-degree phase shift (**2nd from top**) is shifted in the negative direction from the origin. The composite waveform (a summation of all waveforms, **lower left**) is decoded by Fourier transformation into the corresponding positions and amplitudes (**lower right**).

and RF bandwidth. Overall signal amplitude is dependent on spin density and T1 and T2 relaxation events. This produces a spatial domain profile along the direction of the applied gradient (bottom of Fig. 15-9).

Rotation of the FEG direction incrementally (achieved with simultaneous activation of gradient coil pairs) for each repetition time (TR) interval can provide projections through the object as a function of angle, which is very similar conceptually to the data set that is acquired in CT. This data could produce a tomographic image using filtered backprojection techniques. However, because of sensitivity to motion artifacts and the availability of alternate methods to generate MR images, the projection-reconstruction method has largely given way to phase encoding techniques.

Phase Encode Gradient

Position of the spins in the third spatial dimension is determined with a phase encode gradient (PEG), applied before the frequency encode gradient and after the slice encode gradient, along the third perpendicular axis. Phase represents a variation in the starting point of sinusoidal waves, and can be purposefully introduced with the application of a short duration gradient. After the initial localization of the excited protons in the slab of tissue by the SEG, all spins are in phase coherence (they have the same phase). During the application of the PEG, a linear variation in the precessional frequency of the excited spins occurs across the tissue slab along the direction of the gradient. After the PEG is turned off, spin precession reverts to the Larmor frequency, but now phase shifts are introduced, the magnitude of which are dependent on the spatial position relative to the PEG null and the PEG strength. Phase advances for protons in the positive gradient, and phase retards for protons in the negative gradient, while no phase shift occurs for protons at the null. For each TR interval, a specific PEG strength introduces a specific phase change across the FOV. Incremental change in the PEG strength from positive through negative polarity during the image acquisition produces a

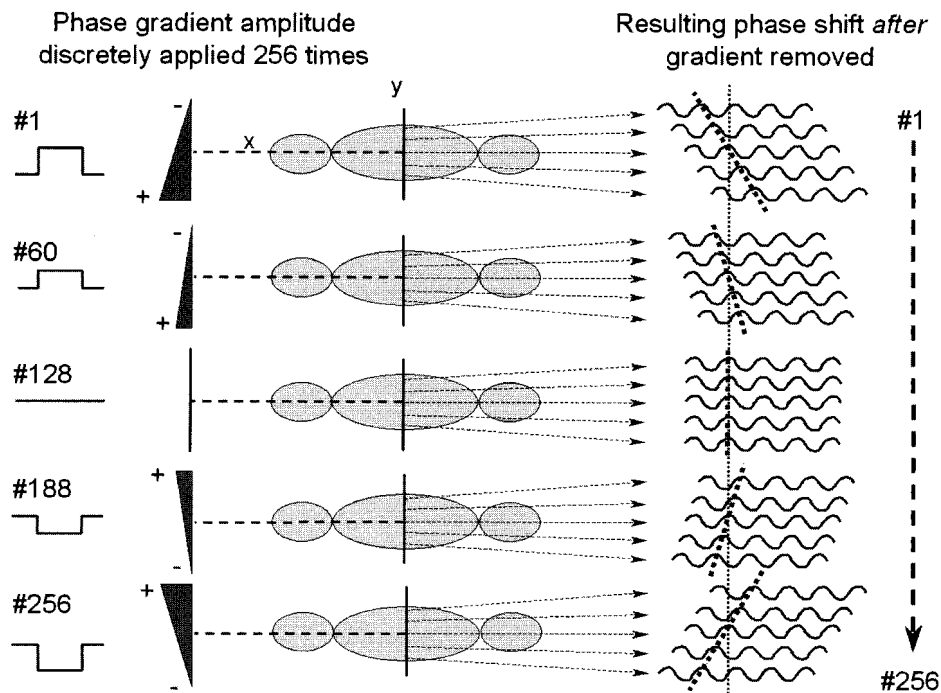


FIGURE 15-11. The phase encode gradient (PEG) produces a spatially dependent variation in angular frequency of the excited spins for a brief duration, and generates a spatially dependent variation in phase when the spins return to the Larmor frequency. Incremental changes in the PEG strength for each TR interval spatially encodes the phase variations: protons at the null of the PEG do not experience any phase change, while protons in the periphery experience a large phase change dependent on the distance from the null. The incremental variation of the PEG strength can be thought of as providing specific "views" of the three-dimensional volume because the SSG and FEG remain fixed throughout the acquisition.

positionally dependent phase shift at each position along the applied phase encode direction (Fig. 15-11). Protons at the center of the FOV (PEG null) do not experience any phase shift. Protons located furthest from the null at the edge of the FOV gain the maximum positive phase shift with the largest positive gradient, no phase shift with the “null” gradient, and maximum negative phase shift with the largest negative gradient. Protons at intermediate distances from the null experience intermediate phase shifts (positive or negative). Thus, each location along the phase encode axis is spatially encoded by the amount of phase shift.

Decoding the spatial position along the phase encode direction occurs by Fourier transformation, only after all of the data for the image have been collected. Symmetry in the frequency domain requires detection of the phase shift direction (positive or negative) to assign correct position in the final image. Since the PEG is incrementally varied throughout the acquisition sequence (e.g., 128 times in a 128×128 MR image), slight changes in position caused by motion will cause a corresponding change in phase, and will be manifested as partial (artifactual) copies of anatomy displaced along the phase encode axis.

Gradient Sequencing

A spin echo sequence is illustrated in Fig. 15-12, showing the timing of the gradients in conjunction with the RF excitation pulses and the data acquisition during the evolution and decay of the echo. This sequence is repeated with slight incremental changes in the phase encode gradient strength to define the three-dimensions in the image.

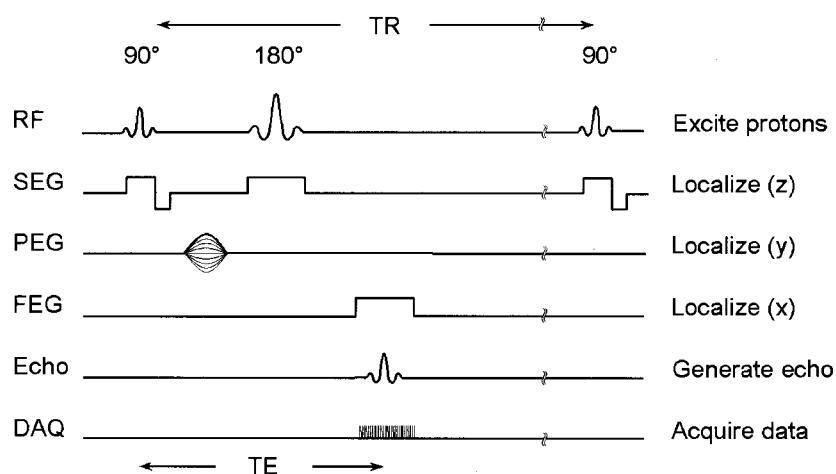


FIGURE 15-12. A typical spin-echo pulse sequence diagram indicates the timing of the SSG, PEG, and FEG during the repetition time (TR) interval, synchronized with the RF pulses and the data acquisition (DAQ) when the echo appears. Each TR interval is repeated with a different PEG strength (this appears as multiple lines in the illustration, but only one PEG strength is applied as indicated by the *bold line* in this example).

15.2 "K-SPACE" DATA ACQUISITION AND IMAGE RECONSTRUCTION

MR data are initially stored in the k-space matrix, the "frequency domain" repository (Fig. 15-13). K-space describes a two-dimensional matrix of positive and negative spatial frequency values, encoded as complex numbers (e.g., $a + bi$, $i = \sqrt{-1}$). The matrix is divided into four quadrants, with the origin at the center representing frequency = 0. Frequency domain data are encoded in the k_x direction by the frequency encode gradient, and in the k_y direction by the phase encode gradient in most image sequences. Other methods for filling k-space include spiral acquisition (Fig. 15-24). The lowest spatial frequency increment (the fundamental frequency) is the bandwidth across each pixel (see Table 15-1). The maximum useful frequency (the Nyquist frequency) is equal to $\frac{1}{2}$ frequency range across the k_x or k_y directions, as the frequencies are encoded from $-f_{\max}$ to $+f_{\max}$. The periodic nature of the frequency domain has a built-in symmetry described by "symmetric" and "antisymmetric" functions (e.g., cosine and sine waves). "Real," "imaginary," and "magnitude" describe specific phase and amplitude characteristics of the composite MR frequency waveforms. Partial acquisitions are possible (e.g., one-half the data acquisition plus one line) with complex-conjugate symmetry filling the remainder of the k-space matrix.

Two-Dimensional Data Acquisition

Frequency data deposited in the k-space matrix, of a simple sine wave of three cycles per unit distance in the frequency encode direction is repeated for each row until k-space is entirely filled (Fig. 15-14). In this example, there is a constant frequency of 3 cycles/distance in the k_x direction and a constant frequency of 0 cycles/distance in the k_y direction. The two-dimensional Fourier transformation (2DFT) yields the corresponding frequency values (positions) and amplitudes as a single output value

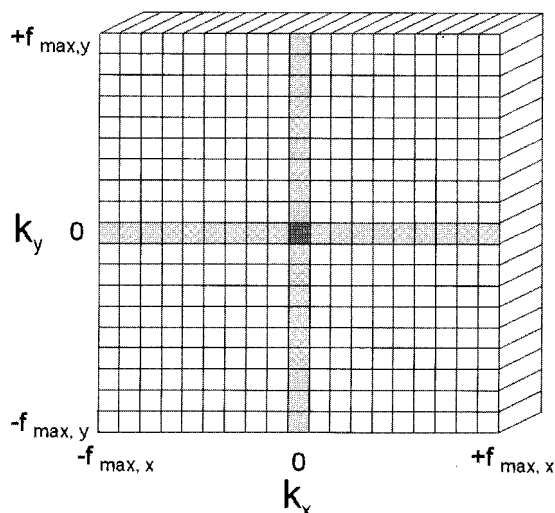


FIGURE 15-13. The k-space matrix is the repository for spatial frequency signals acquired during evolution and decay of the echo. The k_x axis (along the rows) and the k_y axis (along the columns) have units of cycles/unit distance. Each axis is symmetric about the center of k-space, ranging from $-f_{\max}$ to $+f_{\max}$ along the rows and the columns. Low-frequency signals are mapped around the origin of k-space, and high-frequency signals are mapped further from the origin in the periphery. The matrix is filled one row at a time in a conventional acquisition with the FEG induced frequency variations mapped along the k_x axis and the PEG induced phase variations mapped along the k_y axis.

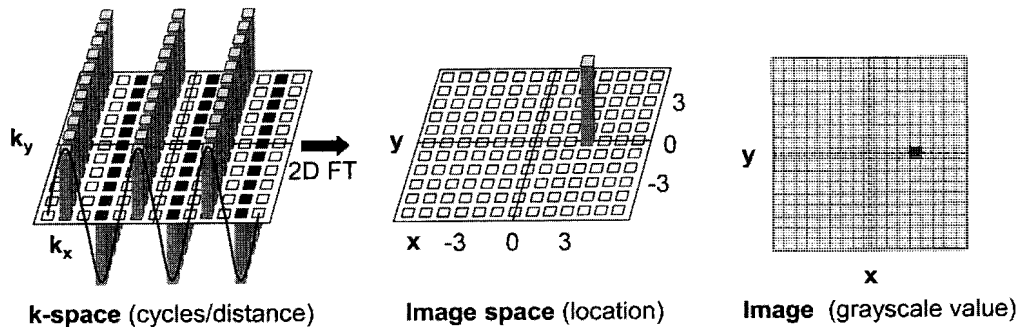


FIGURE 15-14. A simplified example of k-space data acquisition and two-dimensional Fourier transformation has a constant spatial frequency sine wave (3 cycles/distance over the FOV) in the k_x (FEG) direction, and 0 cycles/distance (no phase change over the FOV) in the k_y (PEG) direction. The 2D Fourier transform encodes the variations along the k_x and k_y axes into the corresponding locations along the x and y axes in image space, at $x=3$, $y=0$. Other frequencies and amplitudes, if present, are decoded in the same manner. Grayscale values are determined according to the amplitude of the signal at each spatial frequency. Although not shown, image coordinates are rearranged in the final output image with $x=0$, $y=0$ at the upper left position in the image matrix (see Figs. 15-15 and 15-16).

positioned at position $x = 3$ and $y = 0$ in the spatial domain. The corresponding image represents the amplitude as a gray-scale value.

MR data are acquired as a complex, composite frequency waveform. With methodical variations of the PEG during each excitation, the k-space matrix is filled to produce the desired variations across the frequency and phase encode directions. A summary of the 2D spin echo data acquisition steps follows (the list numbers correspond to the numbers in Fig. 15-15):

1. A narrow band RF excitation pulse is applied simultaneously to the slice select gradient. Energy absorption is dependent on the amplitude and duration of the RF pulse at resonance. The longitudinal magnetization is converted to transverse magnetization, the extent of which is dependent on the saturation of the spins and the angle of excitation. A 90-degree flip angle produces the maximal transverse magnetization.
2. A phase encode gradient is applied for a brief duration to create a phase difference among the spins along the phase encode direction; this produces several "views" of the data along the k_y axis, corresponding to the strength of the PEG.
3. A refocusing 180-degree RF pulse is delivered after a selectable delay time, $TE/2$. This pulse inverts the direction of the individual spins and reestablishes the phase coherence of the transverse magnetization with the formation an echo at a time TE .
4. During the echo formation and subsequent decay, the frequency encode gradient is applied orthogonal to both the slice encode and phase encode gradient directions. This encodes the precessional frequencies spatially along the readout gradient.
5. Simultaneous to the application of the frequency encode gradient and echo formation, the computer acquires the time-domain signal using an analog to digital converter (ADC). The sampling rate is determined by the excitation bandwidth. A one-dimensional Fourier transform converts the digital data into

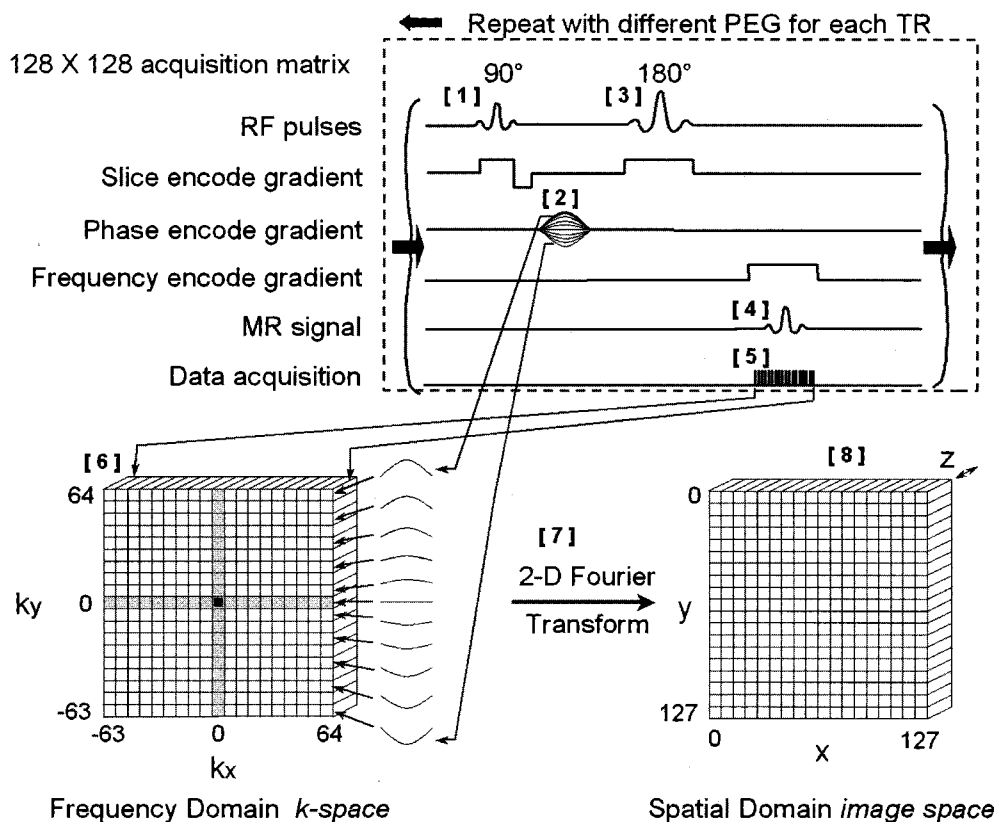


FIGURE 15-15. MR data are acquired into the spatial frequency domain k-space matrix, the repository of spatial frequency signals. Each row in k-space represents spatially dependent frequency variations under a fixed FEG strength, and each column represents spatially dependent phase shift variations under an incrementally varied PEG strength. Data are placed in the k-space matrix in a row determined by the PEG strength for each TR interval. The grayscale image is constructed from the two-dimensional Fourier transformation of the filled k-space matrix, by sequential application of one-dimensional transforms along each row with intermediate results stored in a buffer, and then along each column of the intermediate buffer. The output image matrix is arranged with the image coordinate pair, $x=0$, $y=0$ at the upper left of the image matrix. (The numbers in brackets relate to the descriptions in the text.)

discrete frequency values and corresponding amplitudes. Proton precessional frequencies determine position along the k_x (readout) direction.

6. Data are deposited in the k-space matrix in a row, specifically determined by the phase encode gradient strength applied during the excitation. Incremental variation of the PEG throughout the acquisition fills the matrix one row at a time. In some instances, the phase encode data are acquired in nonsequential order (phase reordering) to fill portions of k-space more pertinent to the requirements of the exam (e.g., in the low-frequency, central area of k-space). After the matrix is filled, the columns contain positionally dependent phase change variations along the k_y (phase encode) direction.
7. The two-dimensional inverse Fourier transform decodes the frequency domain information piecewise along the rows and then along the columns of k-space. Spatial and contrast characteristics of the tissues are manifested in the image.

8. The final image is a spatial representation of the proton density, T1, T2, and flow characteristics of the tissue using a gray-scale range. Each pixel represents a voxel; the thickness is determined by the slice encode gradient strength and RF frequency bandwidth.

The bulk of information representing the lower spatial frequencies is contained in the center of k-space, whereas the higher spatial frequencies are contained in the periphery. A gray-scale representation of k-space data for a sagittal slice of an MR brain acquisition (Fig. 15-16A) shows a concentration of information around the origin (the k-space images are logarithmically amplified for display of the lowest amplitude signals). Two-dimensional Fourier transformation converts the data into a visible image (Fig. 15-16B). Segmenting a radius of 25 pixels out of 128 in the central area (Fig. 15-16C) and zeroing out the periphery extracts a majority of the low-frequency information and produces the corresponding image (Fig. 15-16D); zeroing out the central portion and leaving the

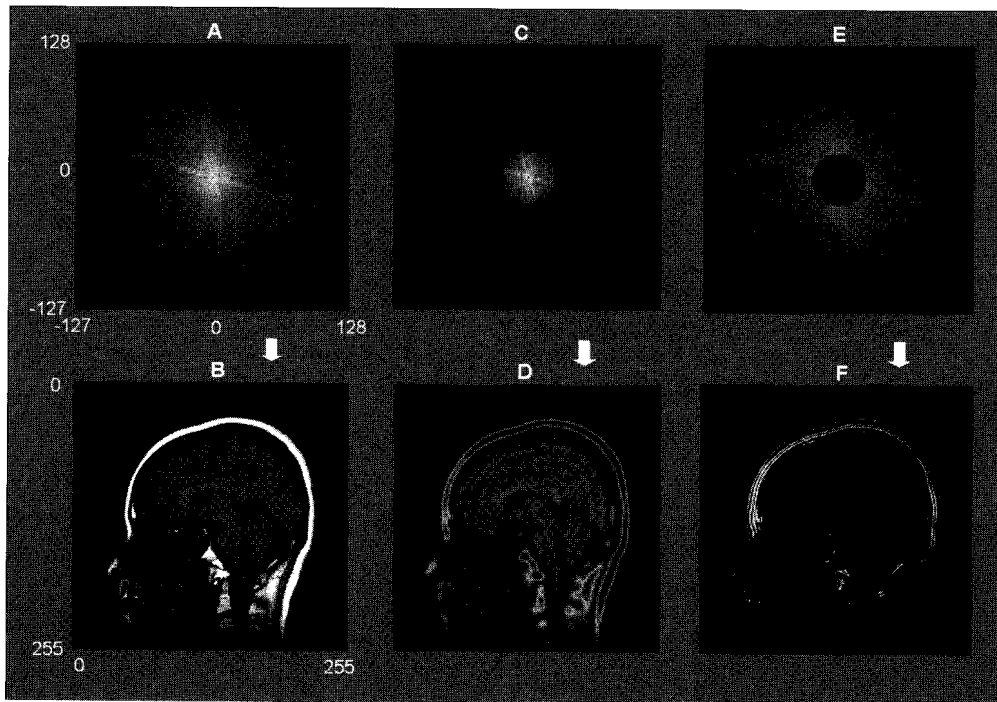


FIGURE 15-16. Image representations of k-space segmentation. The central area of k-space represents the bulk of the anatomy, while the outer areas of the k-space contain the detail and resolution components of the anatomy, as shown by reconstructed images. **A:** Gray-scale-encoded data in a 256×256 matrix (incremental frequency values from -127 to $+128$ in the k_x and k_y directions). Logarithmic amplification is applied to the image display to increase the visibility of the lower amplitude echoes away from the origin. Note the structural symmetry in the data from opposite quadrants. **B:** The 2DFT of the data in **A** produces the spatial domain transform of the full-fidelity MR image. **C:** The inner area of the k-space is segmented out to a radius of 25 pixels from the origin, and zeroed elsewhere. **D:** The 2DFT of the data in **C** shows a blurred, but easily recognized image that uses only the first 25 fundamental frequency values (out of a total of 128 frequency values). **E:** The peripheral area of the k-space is segmented from the inner 25 frequency values around the origin. **F:** The corresponding 2DFT of the data in **E** produces an image of mainly detail and noise.

peripheral areas (Fig. 15-16E) isolates the higher spatial frequency signals (Fig. 15-16F). Clearly, information near the center of k-space provides the large area contrast in the image, while the outer areas in k-space contribute to the resolution and detail.

Two-Dimensional Multiplanar Acquisition

Direct axial, coronal, sagittal, or oblique planes can be obtained by energizing the appropriate gradient coils during the image acquisition. The slice encode gradient determines the orientation of the slices, where axial uses z-axis coils, coronal uses y-axis coils, and sagittal uses x-axis coils for selection of the slice orientation (Fig. 15-17). Oblique plane acquisition depends on a combination of the x-, y-, and z-axis coils energized simultaneously. Phase and frequency encode gradients are perpendicular to the slice encode gradient. Data acquisition into k-space remains the same, with the frequency encode gradient along the k_x axis and the phase encode along the k_y axis.

Acquisition Time, 2DFT Spin Echo Imaging

The time required to acquire an image is equal to

$$TR \times \text{No. of Phase Encode Steps} \times \text{No. of Signal Averages}$$

For instance, in a spin echo sequence for a 256×192 image matrix and two averages per phase encode step with a $TR = 600$ msec, the imaging time is approximately equal to $0.6 \text{ seconds} \times 192 \times 2 = 230.4 \text{ seconds} = 3.84 \text{ minutes}$. In instances where nonsquare pixel matrices are used, e.g., 256×192 , 256×128 , etc., the phase encode direction is typically placed along the small dimension of the image matrix. This reduces the time of acquisition but also sacrifices spatial resolution due to the smaller image matrix. Other clinically useful methods reduce image acquisition time as discussed below.

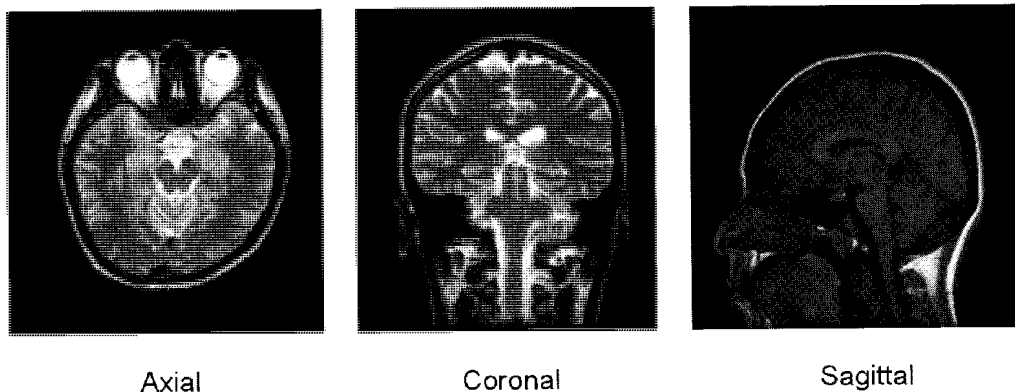


FIGURE 15-17. Direct acquisitions of axial, coronal, and sagittal tomographic images are possible by electronically energizing the magnetic field gradients in a different order without moving the patient. Oblique planes can also be obtained.

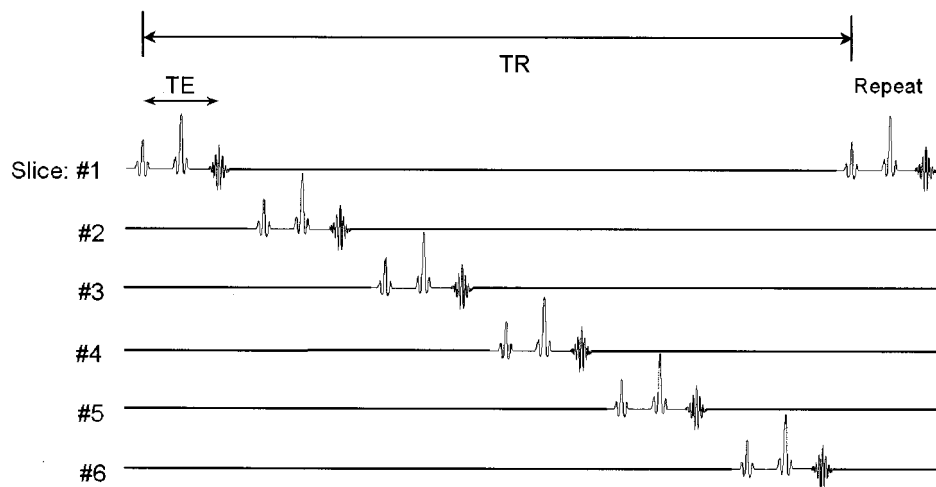


FIGURE 15-18. Multislice two-dimensional image acquisition is accomplished by discretely exciting different slabs of tissue during the TR period; appropriate changes of RF excitation, SEG, PEG, and FEG are necessary. Overlap of excitation volumes is a problem that leads to partial saturation and loss of signal in adjacent slices.

Multislice Data Acquisition

The average acquisition time per slice in a single-slice spin echo sequence is clinically unacceptable. However, the average time per slice is significantly reduced using multiple slice acquisition methods. Several slices within the tissue volume are selectively excited during a TR interval to fully utilize the (dead) time waiting for longitudinal recovery in a specific slice (Fig. 15-18). This requires cycling all of the gradients and tuning the RF excitation pulse many times during the TR interval. The total number of slices is a function of TR, TE, and machine limitations:

$$\text{Total Number of Slices} = \text{TR} / (\text{TE} + C)$$

where C is a constant dependent on the MR equipment capabilities (computer speed, gradient capabilities, RF cycling, etc.). Long TR acquisitions such as proton density and T2-weighted sequences provide a greater number of slices than T1-weighted sequences with a short TR. The chief trade-off is a loss of tissue contrast due to cross-excitation of adjacent slices, causing undesired spin saturation as explained below (see Artifacts).

Data Synthesis

Data “synthesis” takes advantage of the symmetry and redundant characteristics of the frequency domain signals in k-space. The acquisition of as little as one-half the data plus one row of k-space allows the mirroring of “complex-conjugate” data to fill the remainder of the matrix (Fig. 15-19). In the phase encode direction, “half Fourier,” “ $\frac{1}{2}$ NEX,” or “phase conjugate symmetry” (vendor specific names) techniques effectively reduce the number of required TR intervals by one-half plus one line, and thus reduces the acquisition time by nearly one-half. In the frequency encode direction “fractional echo” or “read conjugate symmetry” refers to reading

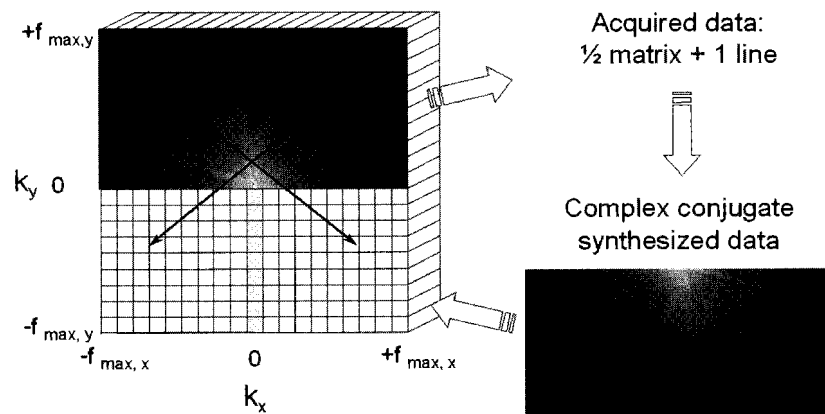


FIGURE 15-19. Data synthesis uses the redundant characteristics of the frequency domain. This is an example of phase conjugate symmetry, in which one-half of the PEG views plus one extra are acquired, and the complex conjugate of the data is reflected in the symmetric quadrants. Acquisition time is thus reduced by approximately one-half, although image noise is increased by approximately $\sqrt{2}$ (~40%).

only a small fraction of the echo. While there is no scan time reduction when all the phase encode steps are acquired, there is a significant echo time reduction, which reduces motion-related artifacts, such as dephasing of blood. The penalty for either half Fourier or fractional echo techniques is a reduction in the SNR (caused by a reduced number of excitations or data sampling in the volume) and the potential for artifacts if the approximations in the complex conjugation of the signals are not accurate. Other inaccuracies result from inhomogeneities of the magnetic field, imperfect linear gradient fields, and the presence of magnetic susceptibility agents in the volume being imaged.

Fast Spin Echo Acquisition

The fast spin echo (FSE) technique uses multiple phase encode steps in conjunction with multiple 180-degree refocusing RF pulses per TR interval to produce a train of up to 16 echoes. Each echo experiences differing amounts of phase encoding that correspond to different lines in k-space (Fig. 15-20). The effective echo time is determined when the central views in k-space are acquired, which are usually the first echoes; subsequent echoes are usually spaced apart with the same echo spacing time, with the latest echoes furthest away from the center. This phase ordering achieves the best SNR by acquiring the low-frequency information with the least T2 decay. The FSE technique has the advantage of spin echo image acquisition, namely immunity from external magnetic field inhomogeneities, with 4× to 8× to 16× faster acquisition time. However, each echo experiences different amounts of T2 decay, which causes image contrast differences when compared with conventional spin echo images of similar TR and TE. Lower signal levels in the later echoes produce less SNR, and fewer images can be acquired in the image volume during the same acquisition. A T2-weighted spin echo image (TR = 2,000 msec, 256 phase encode steps, one average) requires approximately 8.5 minutes, while a

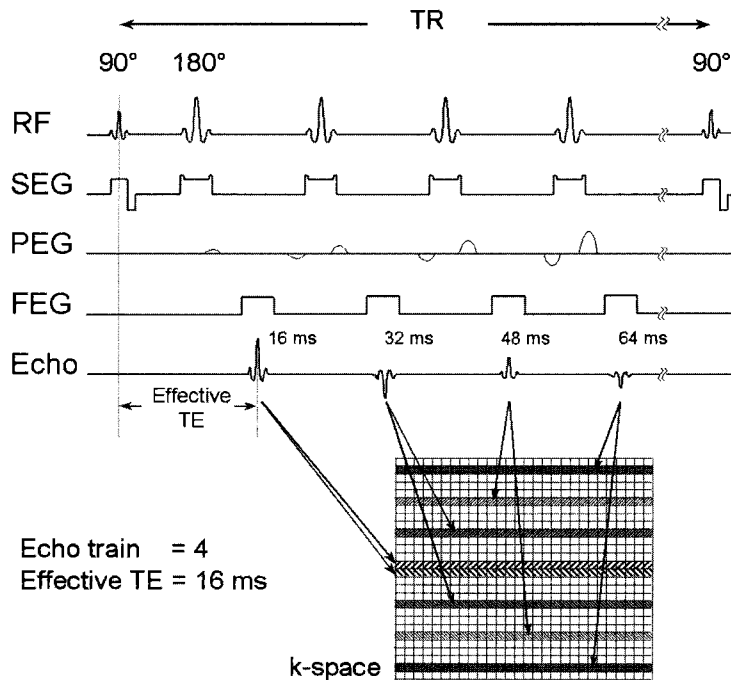


FIGURE 15-20. Conventional fast spin echo (FSE) uses multiple 180-degree refocusing RF pulses per TR interval with incremental changes in the PEG to fill several views in the k-space. This example illustrates an echo train length of four, with an “effective” TE equal to 16 msec (when the central area of the k-space is filled). The reversed polarity PEG steps reestablish coherent phase before the next gradient application. Slightly different PEG strengths are applied to fill adjacent areas during each TR, and this continues until all views are recorded. As shown, data can be mirrored using conjugate symmetry to reduce the overall time by another factor of two.

corresponding FSE with an echo train length of 4 (Fig. 15-20) requires about 2.1 minutes. Specific FSE sequences for T2 weighting and multiecho FSE are employed with variations in phase reordering and data acquisition. FSE is also known as “turbo spin echo,” or “RARE” (rapid acquisition with refocused echoes).

Inversion Recovery Acquisition

The inversion recovery pulse sequence uses an initial 180-degree RF pulse to excite the slice of tissue by inverting the spins. A 90-degree pulse is applied after a delay TI (inversion time) to convert the recovered longitudinal magnetization into the transverse plane. A second 180-degree pulse is applied at TE/2 to invert the spins and produce an echo at a time TE after the 90-degree pulse (Fig. 15-21). During RF excitation, the SSG is applied to localize the spins in the desired plane. Phase and frequency encoding occur similarly to the spin echo pulse sequence. As the TR is typically long (e.g., ~3,000 msec), several slices can be acquired within the volume of interest. STIR (short tau inversion recovery) and FLAIR (fluid attenuated inversion recovery) pulse sequences are commonly used.

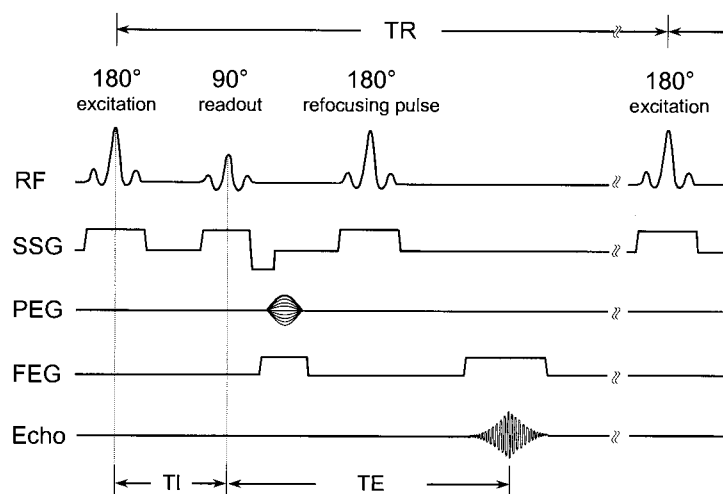


FIGURE 15-21. Inversion recovery pulse sequence uses an initial 180-degree pulse, followed by inversion time (TI) with 90-degree pulse and a second 180-degree pulse at a time TE/2.

Gradient Recalled Echo Acquisition

The gradient recalled echo (GRE) pulse sequence is similar to a standard spin echo sequence with a readout gradient reversal substituting for the 180-degree pulse (Fig. 15-22). Repetition of the acquisition sequence occurs for each PEG step and with each average. With small flip angles and gradient reversals, a considerable reduction in TR and TE is possible for fast image acquisition; however, the ability to acquire multiple slices is compromised. A PEG rewinder pulse of opposite polarity is applied to maintain phase relationships from pulse to pulse.

Gradient echo images require an acquisition time equal to

$$TR \times \text{No. of Phase Encode Steps} \times \text{No. of Signal Averages}$$

the same as the conventional spin echo techniques. For instance, in a gradient-echo sequence for a 256×192 image matrix, two averages, and a TR = 30 msec, the imag-

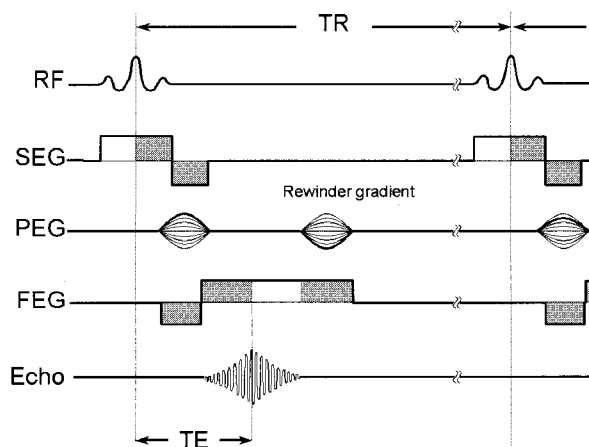


FIGURE 15-22. A conventional gradient recalled echo sequence uses a small flip angle RF pulse simultaneous to the SSG. Phase and frequency encode gradients are applied shortly afterward (with a TE of less than 3 msec in certain sequences). A PEG "rewinder" (reverse polarity) reestablishes the phase conditions prior to the next pulse, simultaneous with the extended FEG duration.

ing time is approximately equal to $192 \times 2 \times 0.03$ seconds = 15.5 seconds. A conventional spin echo requires a time equal to 3.84 minutes for a TR = 600 msec. The compromises of the GRE pulse sequence include SNR losses, magnetic susceptibility artifacts, and less immunity from magnetic field inhomogeneities. In addition, the multiple slice acquisition with the spin echo sequence reduces the effective time per image, and thus the comparison of multiple slice acquisition with GRE imaging time is closer than calculated above, since usually only one slice can be acquired with a very short TR interval. There are several acronyms for gradient recalled echo, including GRASS, FISP, Spoiled GRASS, FLASH, etc.

Echo Planar Image Acquisition

Echo planar image (EPI) acquisition is a technique that provides extremely fast imaging time. Single-shot (all of the image information is acquired within 1 TR interval) or multishot EPI has been implemented. For single-shot EPI, image acquisition typically begins with a standard 90-degree flip, then a PEG/FEG gradient application to initiate the acquisition of data in the periphery of the k-space, followed by a 180-degree echo-producing RF pulse. Immediately after, an oscillating readout gradient and phase encode gradient “blips” are continuously applied to stimulate echo formation and rapidly fill k-space in a stepped “zig-zag” pattern (Fig. 15-23). The

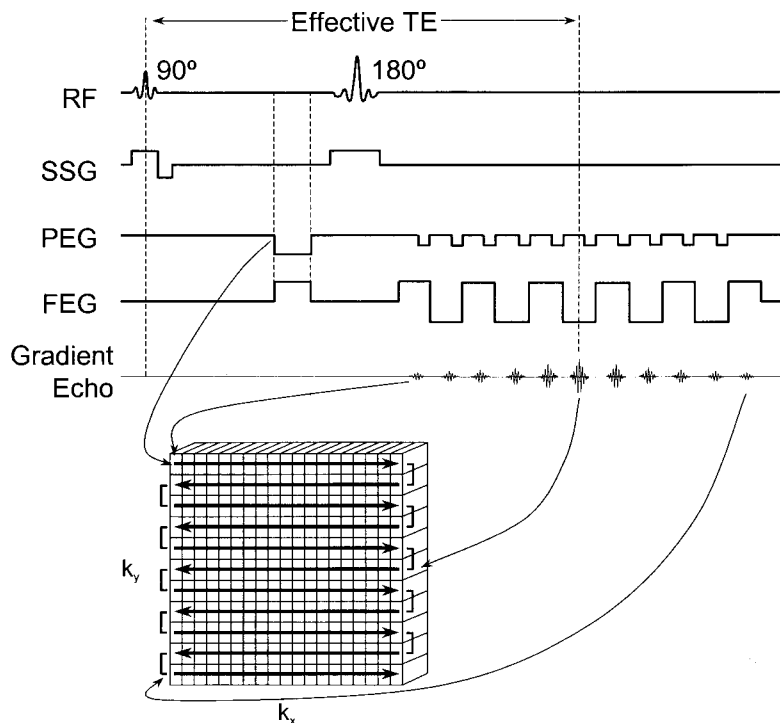


FIGURE 15-23. Single-shot echo planar image (EPI) acquisition sequence. Data are deposited in the k-space with an initial large PEG application to locate the initial row position, followed by phase encode gradient “blips” simultaneous with FEG oscillations to fill the k-space line by line by introducing one-row phase changes in a zigzag pattern. Image matrix sizes of 64×64 and 128×64 are common for EPI acquisitions.

“effective” echo time occurs at a time TE , when the maximum amplitude of the gradient echoes occurs. Acquisition of the data must proceed in a period less than $T2^*$ (around 25 to 50 msec), placing high demands on the sampling rate, the gradient coils (shielded coils are required, with low induced “eddy currents”), the RF transmitter/receiver, and RF energy deposition limitations.

Echo planar images typically have poor SNR, low resolution (matrices of 64×64 or 128×64 are typical), and many artifacts, particularly of chemical shift and magnetic susceptibility origin. Nevertheless, this technique offers real-time “snapshot” image capability with 50 msec or less total acquisition time. EPI is emerging as a clinical tool for studying time-dependent physiologic processes and functional imaging. Multishot echo planar acquisitions with spin echo readout or gradient recalled echo readout are also used.

Other fast image imaging techniques use TR intervals of 3 msec or less with specific gradient coils and high-speed data acquisition capabilities. The extremely fast repetition time provides abilities for cardiac evaluation when used in conjunction with electrocardiogram (ECG) gating for high temporal and reasonable spatial resolution image acquisition.

Spiral K-Space Acquisition

An alternate method of filling the k-space matrix involves the simultaneous oscillation of the PEG and FEG to sample data points during echo formation in a spiral, starting at the origin (the center of the k-space) and spiraling outward to the periphery. The same contrast mechanisms are available in spiral sequences (e.g., T1, T2, spin density weighting), and spin or gradient echoes can be obtained. After acqui-

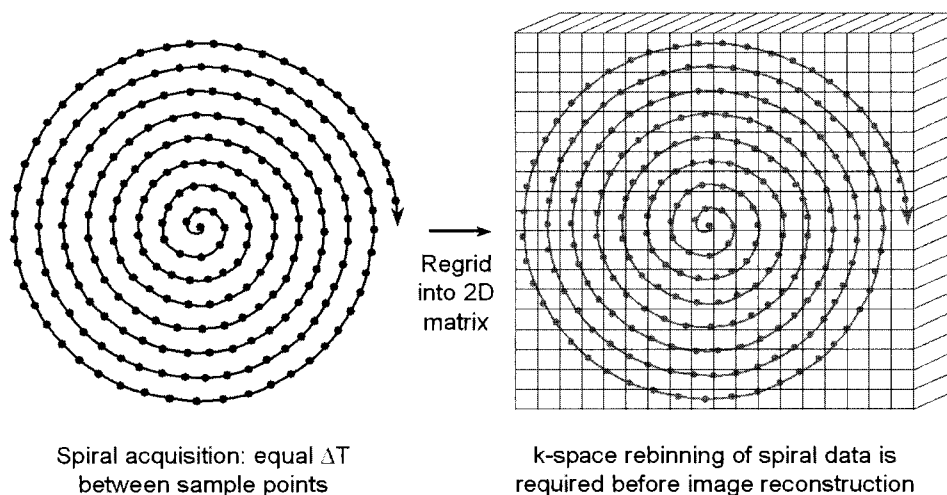


FIGURE 15-24. Spiral data acquisition occurs with sinusoidal oscillation of the x and y gradients 90 degrees out of phase with each other, with samples beginning in the center of the k-space and spiraling out to the periphery. Shown is a single-shot acquisition, and multishot acquisitions (e.g., four interleaved spiral trajectories) are also employed. Image contrast is better preserved for short $T2^*$ decay, since the highest signals are mapped into the central portion of the k-space. Regridding of the data by interpolation into the k_x , k_y matrix is required to apply 2DFT image reconstruction.

sition of the signals (Fig. 15-24), an additional postprocessing step, regridding, is necessary to convert the spiral data into the rectilinear matrix for 2DFT. Spiral scanning is an efficient method for acquiring data and sampling information in the center of k-space, where the bulk of image information is contained. This technique is gaining popularity for a variety of imaging sequences, including cardiac dynamic imaging and functional neuroimaging. Like EPI, spiral scanning is very sensitive to field inhomogeneities and susceptibility agents.

Gradient Moment Nulling

In spin echo and gradient recalled echo imaging, the slice-select and readout gradients are balanced so that the uniform dephasing with the initial gradient application is rephased by an opposite polarity gradient of equal area. However, when moving spins are subjected to the gradients, the amount of phase dispersion is not compensated (Fig. 15-25A). This phase dispersal can cause ghosting (faint, displaced copies of the anatomy) in images. It is possible, however, to rephase the spins by a gradient moment nulling technique. With constant velocity flow (first-order motion), all spins can be rephased using the application of a gradient triplet. In this technique, an initial positive gradient of unit area is followed by a negative gradient of twice the area, which creates phase changes that are compensated by a third positive gradient of unit area. Figure 15-25B depicts the evolution of the spin phase back to zero for both stationary and moving spins. Note that the overall applied gradient has a net area of zero—equal to the sum of the positive and negative areas. Higher-order corrections such as second- or third-order moment nulling to correct for acceleration and other motions are possible, but these techniques require more complicated gradient switching. Moment nulling can be applied to both the slice-select and readout gradients to correct for problems such as motion ghosting as elicited by cerebrospinal fluid (CSF) pulsatile flow.

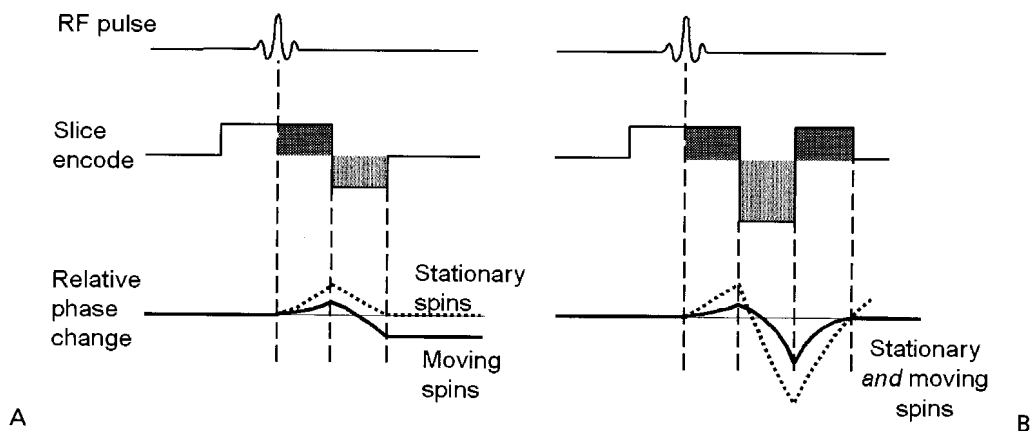


FIGURE 15-25. A: Phase dispersion of stationary and moving spins under the influence of an applied gradient (no flow compensation). The stationary spins at the end of the gradient application return to the original phase state, whereas the moving spins do not. **B:** Gradient moment nulling of first-order linear velocity (flow compensation). In this case, a doubling of the amplitude of the negative gradient is followed by a subsequent positive gradient such that the total summed area is equal to zero. This will return both the stationary spins and the moving spins back to their original phase state.

15.3 THREE-DIMENSIONAL FOURIER TRANSFORM IMAGE ACQUISITION

Three-dimensional image acquisition (volume imaging) requires the use of a broadband, nonselective RF pulse to excite a large volume of spins simultaneously. Two phase gradients are discretely applied in the slice encode and phase encode directions, prior to the frequency encode (readout) gradient (Fig. 15-26). The image acquisition time is equal to

$$\text{TR} \times \text{No. of Phase Encode Steps (z-axis)} \\ \times \text{No. of Phase Encode Steps (y-axis)} \times \text{No. of Signal Averages}$$

A three-dimensional Fourier transform (three 1-D Fourier transforms) is applied for each column, row, and depth axis in the image matrix "cube." Volumes obtained can be either isotropic, the same size in all three directions, or anisotropic, where at least one dimension is different in size. The advantage of the former is equal resolution in all directions; reformations of images from the volume do not suffer from degradations of large sample size. After the spatial domain data (amplitude and contrast) are obtained, individual two-dimensional slices in any arbitrary plane are extracted by interpolation of the cube data.

When using a standard TR of 600 msec with one average for a T1-weighted exam, a $128 \times 128 \times 128$ cube requires 163 minutes or about 2.7 hours! Obviously,

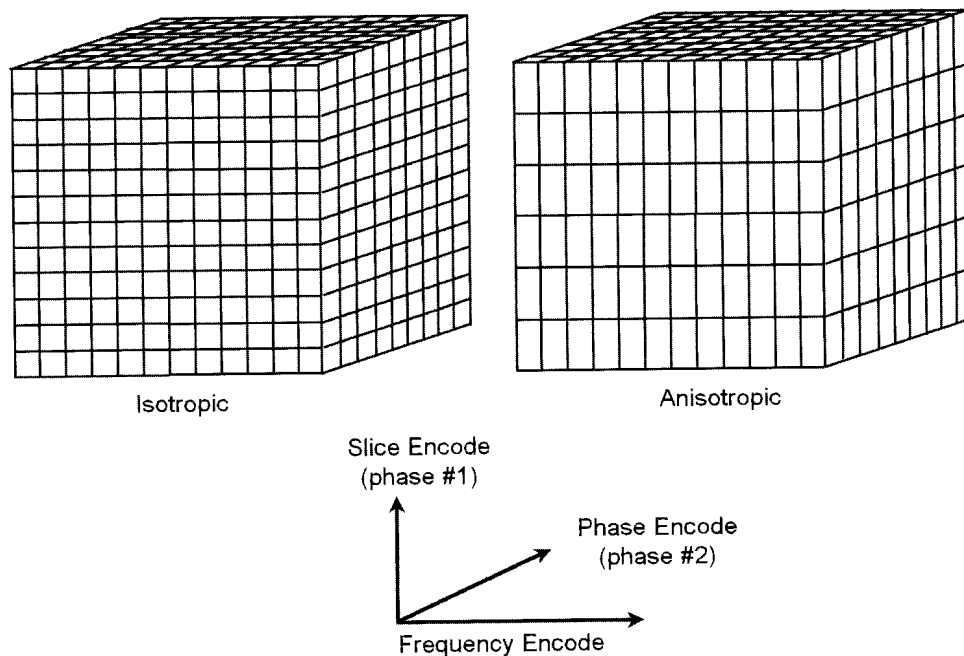


FIGURE 15-26. Three-dimensional image acquisition requires the application of a broadband RF pulse to excite all of the protons in the volume simultaneously, followed by a phase encode gradient along the slice-encode direction, a phase encode gradient along the phase encode direction, and a frequency encode gradient in the readout direction. Spatial location is decoded by a three-dimensional Fourier transform on the three-dimensional k-space data matrix.

this is unacceptable for standard clinical imaging. GRE pulse sequences with TR of 50 msec acquire the same image in about 15 minutes. Another shortcut is with anisotropic voxels, where the phase-encode steps in one dimension are reduced, albeit with a loss of resolution. A major benefit to isotropic three-dimensional acquisition is the uniform resolution in all directions when extracting any two-dimensional image from the matrix cube. In addition, high SNR is achieved compared to a similar two-dimensional image, allowing reconstruction of very thin slices with good detail (less partial volume averaging) and high SNR. A downside is the increased probability of motion artifacts and increased computer hardware requirements for data handling and storage.

15.4 IMAGE CHARACTERISTICS

Spatial Resolution and Contrast Sensitivity

Spatial resolution, contrast sensitivity, and SNR parameters form the basis for evaluating the MR image characteristics. The spatial resolution is dependent on the FOV, which determines pixel size, the gradient field strength, which determines the FOV, the receiver coil characteristics (head coil, body coil, various surface coil designs), the sampling bandwidth, and the image matrix. Common image matrix sizes are 128×128 , 256×128 , 256×192 , and 256×256 , with 512×256 , 512×512 , and 1024×512 becoming prevalent. In general, MR provides spatial resolution approximately equivalent to that of CT, with pixel dimensions on the order of 0.5 to 1.0 mm for a high-contrast object and a reasonably large FOV (>25 cm). A 25-cm FOV and a 256×256 matrix will have a pixel size on the order of 1 mm. In small FOV acquisitions with high gradient strengths and surface coils, the effective pixel size can be smaller than 0.1 to 0.2 mm (of course, the FOV is extremely limited). Slice thickness in MRI is usually 5 to 10 mm and represents the dimension that produces the most partial volume averaging.

It is often assumed that higher magnetic field strengths provide better resolution. This is partially (but not totally) true. A higher field strength magnet generates a larger SNR (see the next subsection) and therefore allows thinner slice acquisition for the same SNR. Improved resolution results from the reduction of partial volume effects. However, with higher magnetic field strengths, increased RF absorption (heating) occurs. Additional compromises of higher field strengths include increased artifact production and a lengthening of T1 relaxation. Higher field strength magnets increase the T1 relaxation constant, which decreases T1 contrast sensitivity because of increased saturation of the longitudinal magnetization.

Contrast sensitivity is the major attribute of MR. The spectacular contrast sensitivity of MR enables the exquisite discrimination of soft tissues and contrast due to blood flow. This sensitivity is achieved through differences in the T1, T2, spin density, and flow velocity characteristics. Contrast which is dependent upon these parameters is achieved through the proper application of pulse sequences, as discussed previously. MR contrast materials, usually susceptibility agents that disrupt the local magnetic field to enhance T2 decay or provide a relaxation mechanism for enhanced T1 decay (e.g., bound water in hydration layers), are becoming important enhancement agents for the differentiation of normal and diseased tissues. The absolute contrast sensitivity of the MR image is ultimately limited by the SNR and presence of image artifacts.

Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) of the MR image is dependent on a number of variables, as included in the equation below for a 2-D image acquisition:

$$\text{SNR} \propto I \times \text{voxel}_{x,y,z} \times \frac{\sqrt{\text{NEX}}}{\sqrt{\text{BW}}} \times f(\text{QF}) \times f(\text{B}) \times f(\text{slice gap}) \times f(\text{reconstruction})$$

where

I = intrinsic signal intensity based on pulse sequence

$\text{voxel}_{x,y,z}$ = voxel volume, determined by FOV, image matrix, and slice thickness

NEX = number of excitations, the repeated signal acquisition into the same voxels, which depends on N_x (# frequency encode data) and N_y (# phase encode steps)

BW = frequency bandwidth of RF transmitter/receiver

$f(\text{QF})$ = function of coil quality factor parameter (tuning the coil)

$f(\text{B})$ = function of magnetic field strength, B

$f(\text{slice gap})$ = function of interslice gap effects, and

$f(\text{reconstruction})$ = function of reconstruction algorithm.

There are numerous dependencies on the ultimate SNR achievable by the MR system. The intrinsic signal intensity based on T1, T2, and spin density parameters has been discussed. Other factors in the equation are explained briefly below.

Voxel Volume

The voxel volume is equal to

$$\text{Volume} = \frac{\text{FOV}_x}{\text{No. of pixels, } x} \times \frac{\text{FOV}_y}{\text{No. of pixels, } y} \times \text{Slice thickness, } z$$

SNR is linearly proportional to the voxel volume. Thus, by reducing the image matrix size from 256×256 to 256×128 , the effective voxel size increases by a factor of two, and therefore increases the SNR by a factor of two for the same image acquisition time (e.g., 256 phase encodes with one average versus 128 phase encodes with two averages).

Signal Averages

Signal averaging (also known as number of excitations, NEX) is achieved by averaging sets of data acquired using an identical pulse sequence. The SNR is proportional to the square root of the number of signal averages. A 2-NEX acquisition requires a doubling (100% increase) of the acquisition time for a 40% increase in the SNR ($\sqrt{2} = 1.4$). Doubling the SNR requires 4 NEX. In some cases, a less than 1 average (e.g., $\frac{1}{2}$ or $\frac{3}{4}$ NEX) can be selected. Here, the number of phase encode steps is reduced by $\frac{1}{2}$ or $\frac{1}{4}$, and the missing data are synthesized in the k-space matrix. Imaging time is therefore reduced by a similar amount; however, a loss of SNR accompanies the shorter imaging times.

RF Bandwidth

The RF coil receiver bandwidth defines the range of frequencies to which the detector is tuned. A narrow bandwidth (a narrow spread of frequencies around the cen-

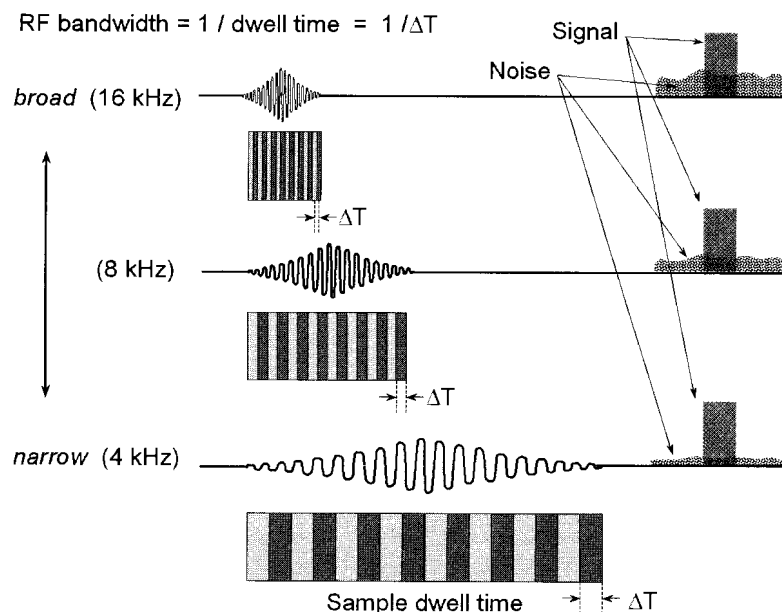


FIGURE 15-27. Evolution of a narrow bandwidth echo and a broad bandwidth echo is determined by the readout gradient strength. A narrow bandwidth is achieved with a low strength gradient, and the echo requires a longer time to evolve. This provides more time to digitize the signal accurately and to smooth out the noise, but an increase in the shortest TE time is required.

ter frequency) provides a higher $\text{SNR} \propto \frac{1}{\sqrt{\text{BW}}}$. A twofold reduction in RF bandwidth—from 8 kHz to 4 kHz, for instance—increases the SNR by $1.4 \times$ (40% increase). Bandwidth is inversely proportional to the sample dwell time, ΔT to sample the signal: $\text{BW} = 1/\Delta T$. Therefore, a narrow bandwidth has a longer dwell time, which averages the random noise variations and reduces the noise relative to the signal (Fig. 15-27), compared to the shorter dwell time for the broad bandwidth signal, where the high-frequency component of the noise is prevalent. The SNR is reduced by the square root of the dwell time. However, any decrease in RF bandwidth must be coupled with a decrease in gradient strength to maintain the same slice thickness, which might be unacceptable if chemical shift artifacts are of concern (see Artifacts, below). Narrower bandwidths also require a longer time for sampling, and therefore affect the minimum TE time that is possible for an imaging sequence.

RF Coil Quality Factor

The coil quality factor is an indication of RF coil sensitivity to induced currents in response to signals emanating in the patient. Coil losses that lead to lower SNR are caused by patient “loading” effects and eddy currents, among other things. Patient loading refers to the electric impedance characteristics of the body, which to a certain extent acts like an antenna. This effect causes a variation in the magnetic field that is different for each patient, and must be measured and corrected for. Consequently, tuning the receiver coil to the resonance frequency is mandatory before

image acquisition. Eddy currents are signals that are opposite of the induced current produced by transverse magnetization in the RF coil, and reduce the overall signal.

The proximity of the receiver coil to the volume of interest affects the coil quality factor, but there are trade-offs with image uniformity. Body receiver coils positioned in the bore of the magnet have a moderate quality factor, whereas surface coils have a high quality factor. With the body coil, the signal is relatively uniform across the FOV; however, with surface coils, the signal falls off abruptly near the edges of the field, limiting the useful imaging depth and resulting in nonuniform brightness across the image.

Magnetic Field Strength

Magnetic field strength influences the SNR of the image by a factor of $B^{1.0}$ to $B^{1.5}$. Thus, one would expect a three- to fivefold improvement in SNR with a 1.5T magnet over a 0.5T magnet. Although the gains in the SNR are real, other considerations mitigate the SNR improvement in the clinical environment, including longer T1 relaxation times and greater RF absorption, as discussed previously.

Cross-Excitation

Cross-excitation occurs from the nonrectangular RF excitation profiles in the spatial domain and the resultant overlap of adjacent slices in multislice image acquisition sequences. This saturates the spins and reduces contrast and the contrast-to-noise ratio. To avoid cross-excitation, interslice gaps or interleaving procedures are necessary (see Artifacts, below).

Image Acquisition and Reconstruction Algorithms

Image acquisition and reconstruction algorithms have a profound effect on SNR. The various acquisition/reconstruction methods that have been used in the past and those used today are, in order of increasing SNR, point acquisition methods, line acquisition methods, two-dimensional Fourier transform acquisition methods, and three-dimensional Fourier transform volume acquisition methods. In each of these techniques, the volume of tissue that is excited is the major contributing factor to improving the SNR and image quality. Reconstruction filters and image processing algorithms will also affect the SNR. High-pass filtration methods that increase edge definition will generally decrease the SNR, while low-pass filtration methods that smooth the image data will generally increase the SNR at the cost of reduced resolution.

15.5 ANGIOGRAPHY AND MAGNETIZATION TRANSFER CONTRAST

Signal from blood is dependent on the relative saturation of the surrounding tissues and the incoming blood flow in the vasculature. In a multislice volume, repeated excitation of the tissues and blood causes a partial saturation of the spins, dependent on the T1 characteristics and the TR of the pulse sequence. Blood outside of

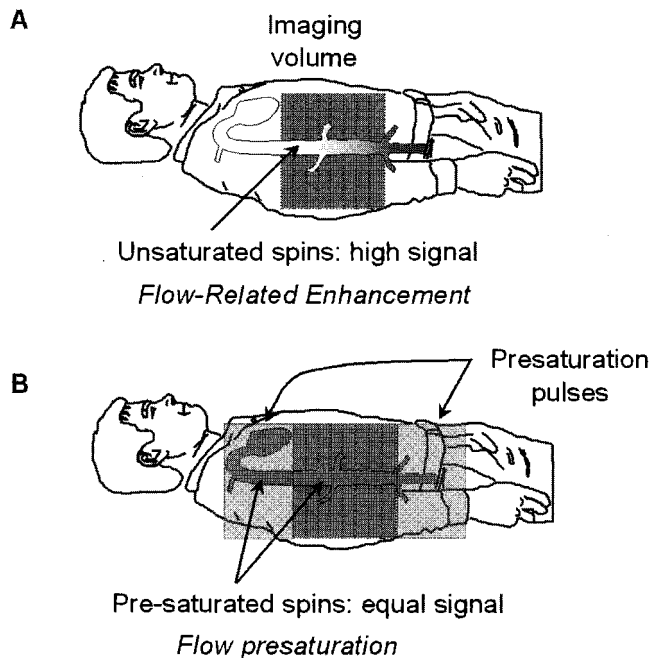


FIGURE 15-28. Presaturation pulses outside of the field of view (FOV) are used to presaturate (B) normally unsaturated spins (A) arriving in the image volume to avoid introducing a variable enhancement signal due to blood flow in a multislice sequence.

the imaged volume does not interact with the RF excitations, and therefore these unsaturated spins may enter the imaged volume and produce a large signal compared to the blood within the volume. This is known as flow-related enhancement. As the pulse sequence continues, the unsaturated blood becomes partially saturated and the spins of the blood produce a similar signal to the tissues in the inner slices of the volume (Fig. 15-28A). In some situations, flow-related enhancement is undesirable and is eliminated with the use of "presaturation" pulses applied to volumes just above and below the imaging volume (Fig. 15-28B). These same saturation pulses are also helpful in reducing motion artifacts caused by adjacent tissues outside the imaging volume.

In some imaging situations, flow-related signal loss occurs when rapidly flowing blood moves through the excited slab of tissues, but does not experience the full refocusing 180-degree pulse, resulting in a flow void, in which the blood appears black in the image.

Exploitation of blood flow enhancement is the basis for MR angiography (MRA). Two techniques to create images of vascular anatomy include time-of-flight and phase contrast angiography.

Time-of-Flight Angiography

The time-of-flight technique relies on the tagging of blood in one region of the body and detecting it in another. This differentiates moving blood from the surround stationary tissues. Tagging is accomplished by spin saturation, inversion, or relaxation to change the longitudinal magnetization of moving blood. The penetration of the tagged blood into a volume depends on the T1, velocity, and direction

of the blood. Since the detectable range is limited by the eventual saturation of the tagged blood, long vessels are difficult to visualize simultaneously in a three-dimensional volume. For these reasons, a two-dimensional stack of slices is typically acquired, where even slowly moving blood can penetrate the region of RF excitation in thin slices (Fig. 15-29A). Each slice is acquired separately, and blood moving in one direction (north or south, e.g., arteries versus veins) can be selected by delivering a presaturation pulse on an adjacent slab superior or inferior to the slab of data acquisition. Thin slices are also helpful in preserving resolution of the flow pattern. Often used for the 2D image acquisition is a "GRASS" or "FISP" GRE technique that produces relatively poor anatomic contrast, yet provides a high-contrast "bright blood" signal.

Two-dimensional MRA images are obtained by projecting the content of the stack of slices at a specific angle through the volume. A maximum intensity projection (MIP) algorithm detects the largest signal along a given ray through the

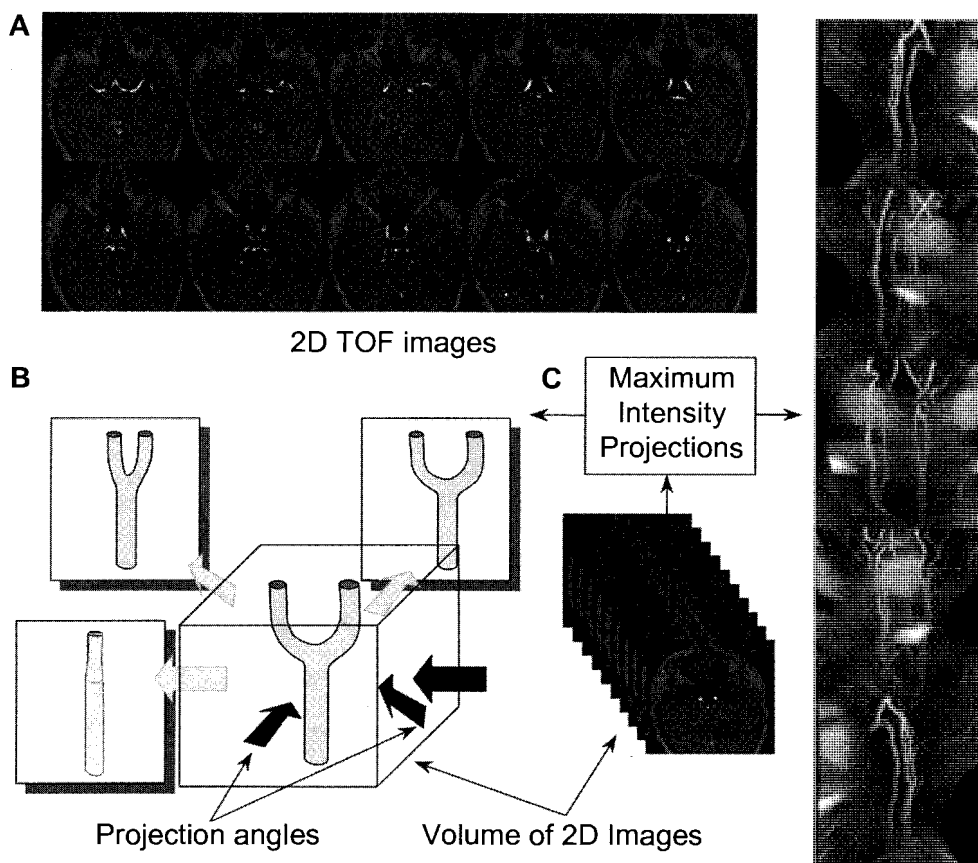


FIGURE 15-29. A: The time of flight MR angiography acquisition collects each slice separately with a sequence to enhance blood flow. A GRASS (FISP) image acquisition pulse sequence is shown. **B:** A simple illustration shows how the maximum intensity projection (MIP) algorithm extracts the high signals in the two-dimensional stack of images, and produces projection images as a function of angle. **C:** MR angiograms of the carotid arteries are projected at several different angles from the two-dimensional stack of flow enhanced time of flight images.

volume and places this value in the image (Fig. 15-29B). The superimposition of residual stationary anatomy often requires further data manipulation to suppress undesirable signals. This is achieved in a variety of ways, the simplest of which is setting a window threshold. Clinical MRA images show the three-dimensional characteristics of the vascular anatomy from several angles around the volume stack (Fig. 15-29C) with some residual signals from the stationary anatomy. Time-of-flight angiography often produces variation in vessel intensity dependent on orientation with respect to the image plane, a situation that is less than optimal.

Phase Contrast Angiography

Phase contrast imaging relies on the phase change that occurs in moving protons such as blood. One method of inducing a phase change is dependent on the application of a bipolar gradient (one gradient with positive polarity followed by a second gradient with negative polarity, separated by a delay time ΔT). In a second acquisition of the same view of the data (same PEG), the polarity of the bipolar gradients is reversed, and moving spins are encoded with negative phase, while the stationary spins exhibit no phase change (Fig. 15-30A). Subtracting the second excitation from the first cancels the magnetization due to stationary spins but enhances magnetization due to moving spins. Alternating the bipolar gradient polarity for each subsequent excitation during the acquisition provides phase contrast image information. The degree of phase shift is directly related to the velocity encoding time, ΔT , between the positive and negative lobes of the bipolar gradients and the velocity of the spins within the excited volume. Proper selection of the velocity encoding time is necessary to avoid phase wrap error (exceeding 180-degree phase change) and to ensure an optimal phase shift range for the velocities encountered. Intensity variations are dependent on the amount of phase shift, where the brightest pixel values represent the largest forward (or backward) velocity, a mid-scale value represents 0 velocity, and the darkest pixel values represent the largest backward (or forward) velocity. Figure 15-30B shows a representative magnitude and phase contrast image of the cardiac vasculature. Unlike the time-of-flight methods, the phase contrast image is inherently quantitative, and when calibrated carefully, provides an estimate of the mean blood flow velocity and direction. Two- and three-dimensional phase contrast image acquisitions for MRA are possible.

Magnetization Transfer Contrast

Magnetization transfer contrast pulse sequences are often used in conjunction with MRA time-of-flight methods. Hydrogen constitutes a large fraction of macromolecules such as proteins. The hydrogen atoms are tightly bound to the macromolecule and have a very short T2 decay with a broad range of resonance frequencies compared to hydrogen in free water. By using an off-resonance RF pulse of $\sim 1,500$ Hz from the Larmor frequency, these protons are excited and become saturated. Water bound to these molecules are influenced by and become partially saturated (see Chapter 14). The outcome is that signal from highly proteinaceous tissues such as brain, liver, and muscles is suppressed, and the contrast of the flow-enhanced signals is increased.

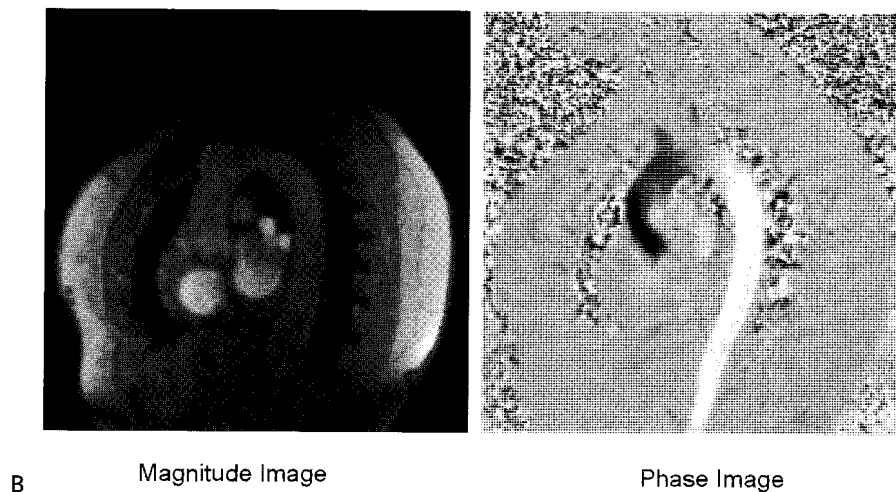
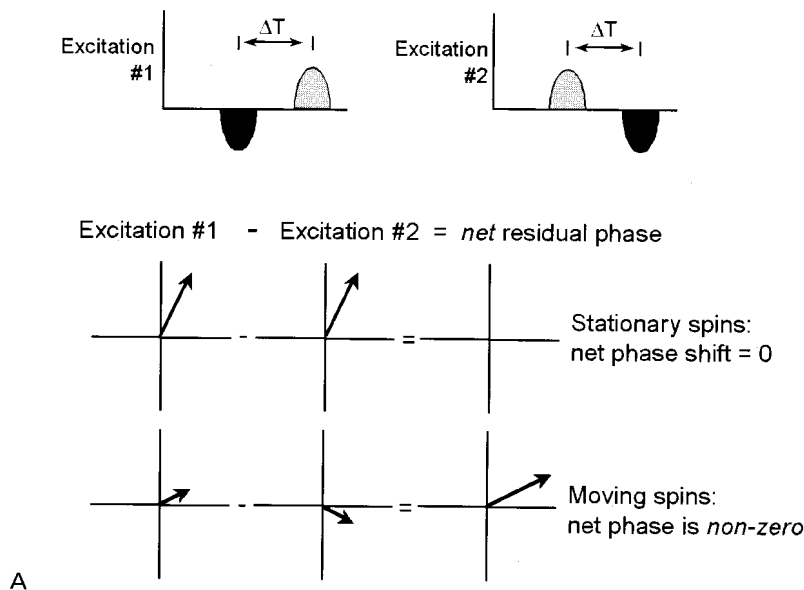


FIGURE 15-30. A: Phase contrast techniques rely on the variable phase change occurring to moving spins when experiencing a “bipolar” gradient compared to no phase change for stationary tissues. By reversing the polarity of the bipolar gradient after each excitation, a subtraction of the alternate excitations will cancel stationary tissue magnetization and will enhance phase differences caused by velocity of moving blood. **B:** Magnitude (*left*) and phase (*right*) images provide contrast of flowing blood. Magnitude images are sensitive to flow, but not to direction; phase images provide direction and velocity information. The blood flow from the heart shows reverse flow in the ascending aorta (*dark area*) and forward flow in the descending aorta at this point in the heart cycle for the phase image. Gray-scale amplitude is proportional to velocity, where intermediate gray-scale indicates no motion. (Images courtesy of Dr. Michael H. Buonocore, MD, PhD, University of California Davis.)

15.6 ARTIFACTS

In MRI, artifacts manifest as positive or negative signal intensities that do not accurately represent the imaged anatomy. Although some artifacts are relatively insignificant and are easily identified, others can limit the diagnostic potential of the exam by obscuring or mimicking pathologic processes or anatomy. One must realize the impact of acquisition protocols and understand the etiology of artifact production to exploit the information they convey.

To minimize the impact of MRI artifacts, a working knowledge of MR physics as well as image acquisition techniques is required. On the one hand, there are many variables and options available that complicate the decision-making process for MR image acquisition. On the other, the wealth of choices enhances the goal of achieving diagnostically accurate images. MRI artifacts are classified into three broad areas—those based on the machine, on the patient, and on signal processing.

Machine-Dependent Artifacts

Magnetic field inhomogeneities are either global or focal field perturbations that lead to the mismapping of tissues within the image, and cause more rapid T2 relaxation. Distortion or misplacement of anatomy occurs when the magnetic field is not completely homogeneous. Proper site planning, self-shielded magnets, automatic shimming, and preventive maintenance procedures help to reduce inhomogeneities. The use of gradient refocused echo acquisition places heightened demands on field uniformity.

Focal field inhomogeneities arise from many causes. Ferromagnetic objects in or on the patient (e.g., makeup, metallic implants, prostheses, surgical clips, dentures) produce field distortions and cause protons to precess at frequencies different from the Larmor frequency in the local area. Incorrect proton mapping, displacement, and appearance as a signal void with a peripherally enhanced rim of increased signal are common findings. Geometric distortion of surrounding tissue is also usually evident. Even nonferromagnetic conducting materials (e.g., aluminum) produce field distortions that disturb the local magnetic environment. Partial compensation by the spin echo (180-degree RF) pulse sequence reduces these artifacts; on the other hand, the gradient refocused echo sequence accentuates distortions, since the protons always experience the same direction of the focal magnetic inhomogeneities within the patient.

Susceptibility Artifacts

Magnetic susceptibility is the ratio of the induced internal magnetization in a tissue to the external magnetic field. As long as the magnetic susceptibility of the tissues being imaged is relatively unchanged across the field of view, then the magnetic field will remain uniform. Any drastic changes in the magnetic susceptibility will distort the magnetic field. The most common susceptibility changes occur at tissue-air interfaces (e.g., lungs and sinuses), which cause a signal loss due to more rapid dephasing (T2*) at the tissue-air interface. Any metal (ferrous or not) may have a significant effect on the adjacent local tissues due to changes in susceptibility and the resultant magnetic field distortions. Paramagnetic agents exhibit a weak magnetization and increase the local magnetic field causing an artifactual reduction in the surrounding T2* relaxation.

Magnetic susceptibility can be quite helpful in some diagnoses. Most notable is the ability to diagnose the age of a hemorrhage based on the signal characteristics of the blood degradation products, which are different in the acute, subacute, and chronic phases. Some of the iron-containing compounds (deoxyhemoglobin, methemoglobin, hemosiderin, and ferritin) can dramatically shorten T1 and T2 relaxation of nearby protons. The amount of associated free water, the type and structure of the iron-containing molecules, the distribution (intracellular versus extracellular), and the magnetic field strength all influence the degree of relaxation effect that may be seen. For example, in the acute stage, T2 shortening occurs due to the paramagnetic susceptibility of the organized deoxyhemoglobin in the local area, without any large effect on the T1 relaxation time. When red blood cells lyse during the subacute stage, the hemoglobin is altered into methemoglobin, and spin-lattice relaxation is enhanced with the formation of a hydration layer, which shortens T1 relaxation, leading to a much stronger signal on T1-weighted images. Increased signal intensity on T1-weighted images not found in the acute stage of hemorrhage identifies the subacute stage. In the chronic stage, hemosiderin, found in the phagocytic cells in sites of previous hemorrhage, disrupts the local magnetic homogeneity, causes loss of signal intensity, and leads to signal void, producing a characteristic dark rim around the hemorrhage site.

Gadolinium-based contrast agents (paramagnetic characteristics shorten T2, and hydration layer interactions shorten T1) are widely used in MRI. Tissues that uptake gadolinium contrast agents exhibit shortened T1 relaxation and demonstrate increased signal on T1-weighted images. Although focal inhomogeneities are generally considered problematic, with certain pathology or anatomy, unique diagnostic information can sometimes be obtained.

Gradient Field Artifacts

Magnetic field gradients spatially encode the location of the signals emanating from excited protons within the volume being imaged. Proper reconstruction requires linear, matched, and properly sequenced gradients. The slice encode gradient defines the volume (slice). Phase and frequency encoding gradients provide the spatial information in the other two dimensions.

Since the reconstruction algorithm assumes ideal, linear gradients, any deviation or temporal instability will be represented as a distortion. A tendency of lower field strength occurs at the periphery of the FOV. Consequently, anatomic compression occurs, especially pronounced on coronal and sagittal images having a large FOV, typically greater than 35 cm (Fig. 15-31). Minimizing the spatial distortion entails either reducing the FOV by lowering the gradient field strength or by holding the gradient field strength and number of samples constant while decreasing the frequency bandwidth.

Anatomic proportions may simulate abnormalities; so square pixel dimensions are necessary. However, if the strength of the frequency encode gradient and the largest phase encode gradient are different, the height or width of the pixels become distorted and can produce nonsquare areas. Ideally, the phase and frequency encoding gradients should be assigned to the smaller and larger dimensions of the object, respectively, to preserve spatial resolution while limiting the number of phase encoding steps. In practice, this is not always possible, because motion artifacts or high-intensity signals that need to be displaced away from important areas of interest after an initial scan might require swapping the frequency and phase encode gradient directions.

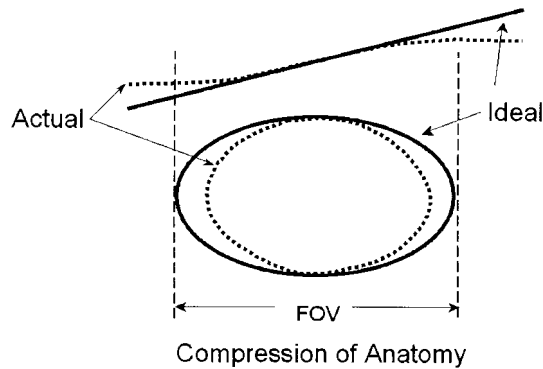


FIGURE 15-31. Magnetic field distortion. Image distortions can often be caused by gradient nonlinearities or gradient failure. In the above example, the strength at the periphery is less than the ideal, resulting in an artifactual compression of the anatomy (*dotted line*). This causes the image to have a “barrel distortion” appearance.

Radiofrequency Coil Artifacts

Surface coils produce variations in uniformity across the image caused by RF attenuation, RF mismatching, and sensitivity falloff with distance. Close to the surface coil, the signals are very strong, resulting in a very intense image signal. Further from the coil, the signal strength drops due to attenuation, resulting in shading and loss of image brightness.

Other common artifacts from RF coils occur with RF quadrature coils (coils that simultaneously measure the signal from orthogonal views) that have two separate amplifier and gain controls. If the amplifiers are imbalanced, a bright spot in the center of the image, known as a center-point artifact, arises as a “0 frequency” direct current (DC) offset. Variations in gain between the quadrature coils can cause ghosting of objects diagonally in the image.

Radiofrequency Artifacts

RF pulses and precessional frequencies of MRI instruments occupy the same frequencies of common RF sources, such as TV and radio broadcasts, electric motors, fluorescent lights, and computers. Stray RF signals that propagate to the MRI antenna can produce various artifacts in the image. Narrow-band noise creates noise patterns perpendicular to the frequency encoding direction. The exact position and spatial extent depends on the resonant frequency of the imager, applied gradient field strength, and bandwidth of the noise. A narrow band pattern of black/white alternating noise produces a zipper artifact. Broadband RF noise disrupts the image over a much larger area of the reconstructed image with diffuse, contrast-reducing “herringbone” artifacts. Appropriate site planning and the use of properly installed RF shielding materials (e.g., a Faraday cage) reduce stray RF interference to an acceptably low level.

RF energy received by adjacent slices during a multislice acquisition due to nonrectangular RF pulses excite and saturate protons in adjacent slices. On T2-weighted images, the slice-to-slice interference degrades the SNR, while on T1-weighted images the extra spin saturation reduces image contrast by reducing longitudinal recovery during the TR interval. Figure 15-32A illustrates a typical truncated sinc RF profile and overlap areas in adjacent slices. Interslice gaps (Fig. 15-32B) reduce the overlap of the profile tails, and pseudorectangular RF pulse profiles reduce the spatial extent of the tails. However, important anatomic findings might exist within the gaps. Slice interleaving is an acquisition technique that can

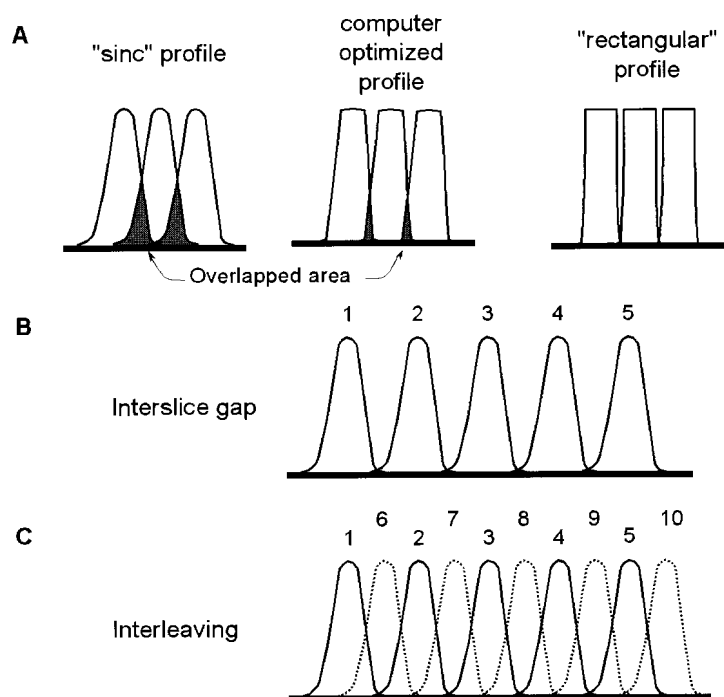


FIGURE 15-32. A: RF profiles and cross-excitation. Poor pulse profiles are caused by RF sinc pulses that are truncated to achieve short pulses. Overlap of the profiles cause unwanted saturation of the signals in adjacent slices, with a loss of SNR and image contrast. Optimized pulses are produced by considering the trade-off of pulse duration versus excitation profile. **B:** Reduction of cross-excitation is achieved with interslice gaps, but anatomy at the gap location is missed. **C:** An interleaving technique acquires the first half of the images with an interslice gap, and the second half of the images positioned in the gaps of the first images. The separation in time reduces the amount of contrast reducing saturation of the adjacent slices.

mitigate cross-excitation by reordering slices into two groups with gaps. During the first half of the TR, the first slices are acquired (Fig. 15-32C, slices 1 to 5), followed by the second group of slices that are positioned in the gap of the first group (Fig. 15-32C, slices 6 to 10). This method reduces cross-excitation by separating the adjacent slice excitations in time. The most effective method lengthens image time by a factor of 2, but achieves “gapless” imaging by acquiring two independent sets of gapped multislice images. The most appropriate solution is to devise RF pulses that approximate a rectangular profile; however, the additional time necessary for producing such an RF pulse can be prohibitive.

K-Space Errors

Errors in k-space encoding affect the reconstructed image, and cause the artifactual superimposition of wave patterns across the FOV. Even one bad single pixel introduces a significant artifact, rendering the image useless (Fig. 15-33). If the bad pixels in the k-space are identified, simple averaging of the signals in adjacent pixels can significantly reduce the artifacts.

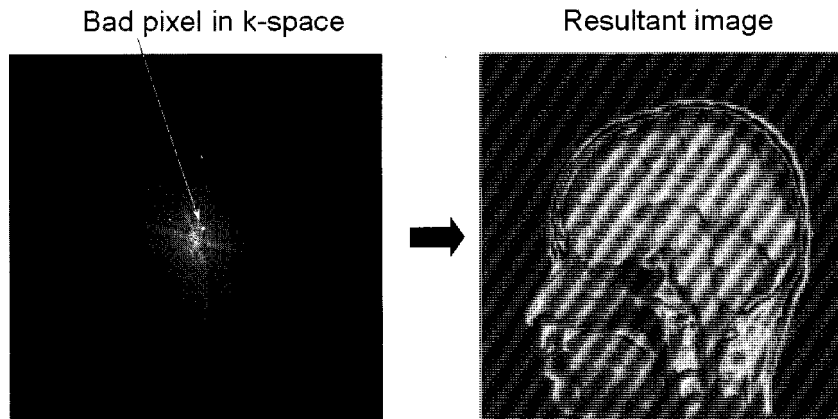


FIGURE 15-33. A single bad pixel in the k-space causes a significant artifact in the reconstructed image. The bad pixel is located at $k_x = 2$, $k_y = 3$, which produces a superimposed sinusoidal wave on the spatial domain image.

Motion Artifacts

The most ubiquitous and noticeable artifacts in MRI arise with patient motion, including voluntary and involuntary movement, and flow (blood, CSF). Although motion artifacts are not unique to MRI, the long acquisition time of certain MRI sequences increases the probability of motion blurring and contrast resolution losses. Motion artifacts mostly occur along the phase encode direction, since adjacent lines of phase-encoded protons are separated by a TR interval that can last 3,000 msec or longer. Even very slight motion will cause a change in the recorded phase variation across the FOV throughout the MR acquisition sequence (Fig. 15-34). The frequency encode direction is less affected, especially by periodic motion,

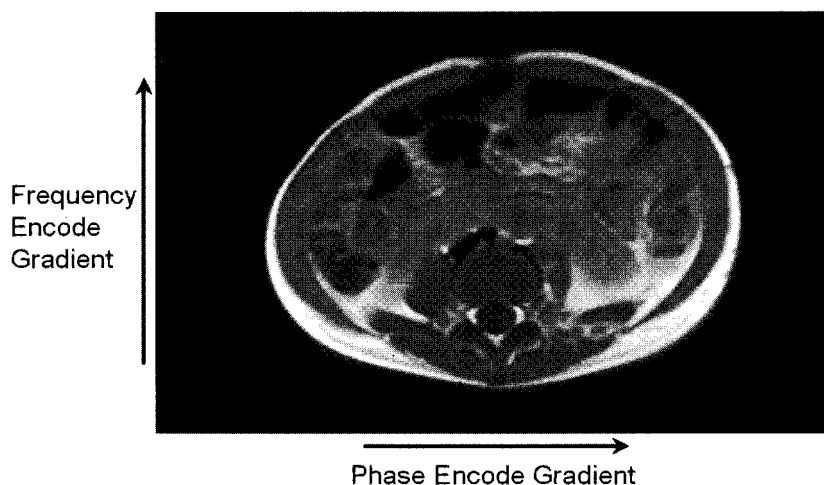


FIGURE 15-34. Motion artifacts, particularly of flow patterns, are most always displayed in the phase encode gradient direction. Slight changes in phase produce multiple ghost images of the anatomy, since the variation in phase caused by motion can be substantial between excitations.

since the evolution of the echo signal, frequency encoding, and sampling occur simultaneously over several milliseconds as opposed to seconds to minutes in the phase direction. Ghost images, which are faint copies of the image displaced along the phase encode direction, are the visual result of patient motion.

Several techniques can compensate for motion-related artifacts. The simplest technique transposes the PEG and FEG to relocate the motion artifacts out of the region of diagnostic interest with the same pulse sequences. This does not reduce the magnitude of the artifacts, however.

There are other motion compensation methods:

1. Cardiac and respiratory gating—signal acquisition at a particular cyclic location synchronizes the phase changes applied across the anatomy (Fig. 15-35).
2. Respiratory ordering of the phase encoding projections based on location within the respiratory cycle.
3. Signal averaging to reduce artifacts of random motion by making displaced signals less conspicuous relative to stationary anatomy.
4. Short TE spin echo sequences (limited to proton density, T1-weighted scans). Long TE scans (T2 weighting) are more susceptible to motion.
5. Gradient moment nulling (additional gradient pulses for flow compensation) to help rephase spins that are dephased due to motion. Most often, these techniques require a longer TE, and are more useful for T2-weighted scans (Fig. 15-25).
6. Presaturation pulses applied outside the imaging region to reduce flow artifacts from inflowing spins (Fig. 15-28).

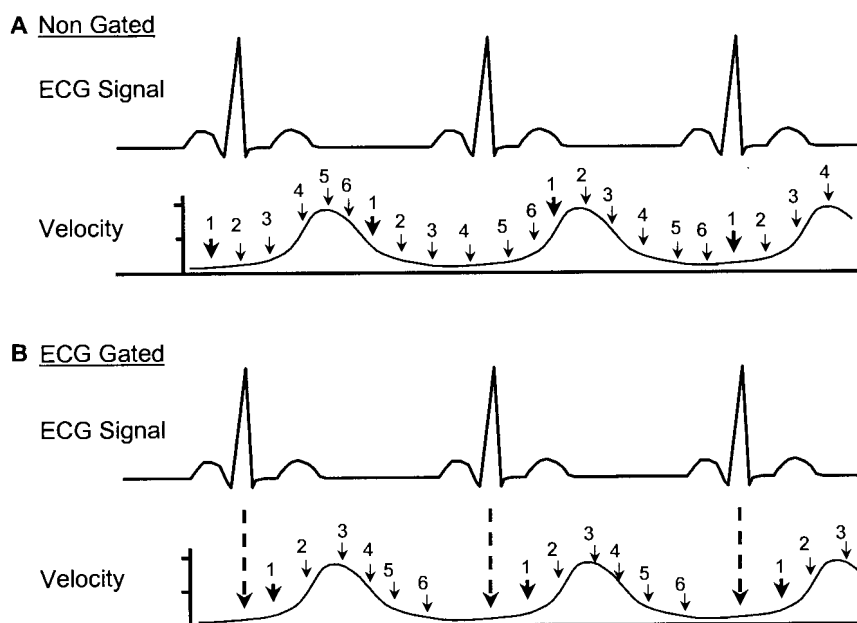


FIGURE 15-35. A: Motion artifacts occur because the multislice image data (indicated as 1 through 6) are not synchronized with the periodicity of the beating heart and the velocity pattern. **B:** Electrocardiogram gated acquisitions synchronize the acquisition of the image data with the R wave (peak wave of ECG signal) to reduce vascular motion artifacts; a reduced number of images or extended acquisition time is required.

On T1-weighted images of the head, blood flow and CSF pulsations are the source of most motion artifacts. Pulse sequences using a gradient moment nulling technique “refocus” moving spins by using switched readout gradient polarities, and these techniques can correct for periodic motion occurring during the TE period. Similarly, on T2-weighted head images the most detrimental motion degradations arise from pulsating CSF. Gradient moment nulling techniques can also be applied to null the signal from these moving spins and improve T2 contrast. Compensation can be applied to one, two, or three axes, and corrections can be first order (linear velocity), second order (acceleration), or higher order, depending on the applied readout gradient waveform. ECG pulse and respiratory gating is a complementary technique that significantly reduces cardiac, CSF, and respiratory motion. The trade-offs of gating are increased image acquisition and operator setup time.

Chemical Shift Artifacts

Chemical shift refers to the resonance frequency variations resulting from intrinsic magnetic shielding of anatomic structures. Molecular structure and electron orbital characteristics produce fields that shield the main magnetic field and give rise to distinct peaks in the MR spectrum. In the case of proton spectra, peaks correspond to water and fat, and in the case of breast imaging, silicone material. Lower frequencies of about 3.5 parts per million for protons in fat and 5.0 parts per million for protons in silicone occur, compared to the resonance frequency of protons in water (Fig. 15-36A). Since resonance frequency increases linearly with field strength, the absolute difference between the fat and water resonance also increases, making high field strength magnets more susceptible to chemical shift artifact. A sample calculation in the example below demonstrates frequency variations in fat and water for two different magnetic field and gradient field strengths.

Data acquisition methods cannot directly discriminate a frequency shift due to the application of a frequency encode gradient or to a chemical shift artifact. Water and fat differences therefore cannot be distinguished by the frequency difference induced by the gradient. The protons in fat resonate at a slightly lower frequency than the corresponding protons in water, and cause a shift in the anatomy (misregistration of water and fat moieties) along to the frequency encode gradient direction (Fig. 15-36B,C).

Example: Numerical calculation of the chemical shift artifact

Field strength:

A 3 parts per million (ppm) chemical shift of water and fat is equivalent to a frequency difference at:

0.5 T of $21 \text{ MHz} \times 3 \text{ ppm} = 63 \text{ Hz}$

1.5 T of $64 \text{ MHz} \times 3 \text{ ppm} = 192 \text{ Hz}$

Thus, the chemical shift is more severe for higher field strength magnets.

Gradient strength, 25 cm (0.25 m) FOV, 256×256 matrix:

A gradient strength of $2.5 \text{ mT/m} \times 0.25 \text{ m} = 0.000625 \text{ T}$

Frequency difference: $0.000625 \text{ T} \times 42.58 \text{ MHz/T} = 26.6 \text{ kHz}/256 \text{ pixels}$
 $= 104 \text{ Hz/pixel}$.

A gradient strength of $10 \text{ mT/m} \times 0.25 \text{ m} = 0.0025 \text{ T}$

Frequency difference: $0.0025 \text{ T} \times 42.58 \text{ MHz/T} = 106.5 \text{ kHz}/256 \text{ pixels}$
 $= 416 \text{ Hz/pixel}$.

Thus, a chemical shift is more severe for lower gradient strengths.

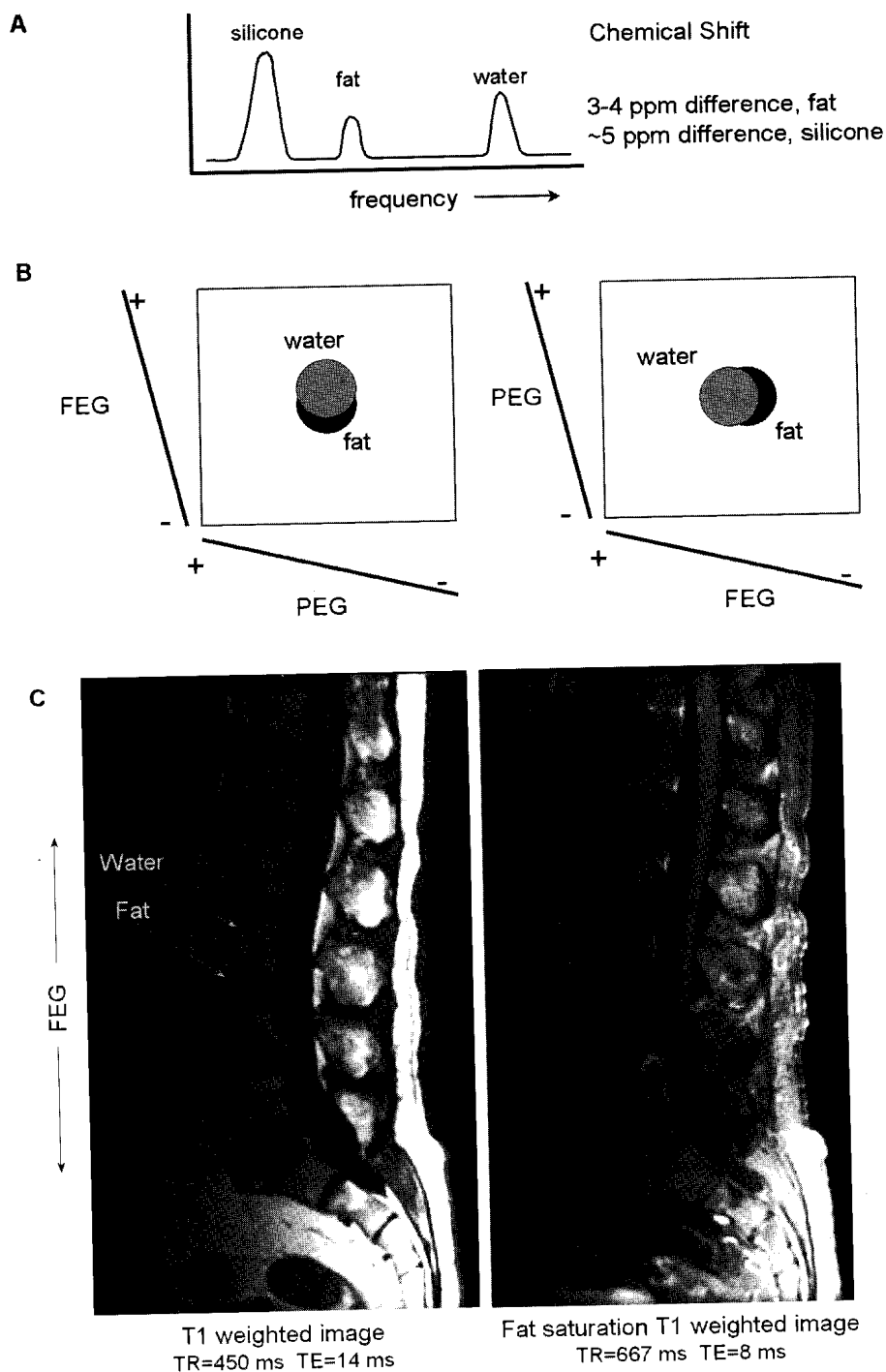


FIGURE 15-36. A: Chemical shift refers to the slightly different precessional frequencies of protons in different materials or tissues. The shifts (in parts per million, ppm) are referenced to water for fat and silicone. **B:** Fat chemical shift artifacts are represented by a shift of water and fat in the images of anatomic structure, mainly in the frequency encode gradient direction. Swapping the phase encode gradient (PEG) and the frequency encode gradient (FEG) will cause a shift of the fat and water components of the tissues. **C:** The *left* image is T1 weighted with bright fat signal and chemical shift artifacts in the water/fat components of the vertebral bodies and other anatomy. The *right* image is T1 weighted with chemical fat saturation pulses to saturate and reduce the fat signal, and to eliminate fat/water chemical shift artifacts. In both images, the FEG is vertically oriented.

A higher frequency encode gradient strength has a larger bandwidth per pixel, and a weaker gradient strength has a smaller bandwidth per pixel. Therefore, with a larger gradient strength, water and fat are contained within the pixel boundaries, while for a lower gradient strength, water and fat are separated by one or more pixels.

RF bandwidth/gradient strength considerations and pulse sequences can mitigate chemical shift problems. Large gradient strength confines the chemical shift within the pixel boundaries, but a significant SNR penalty accrues with the broad RF bandwidth required for a given slice thickness. Instead, lower gradient strengths and narrow bandwidths can be used in combination with off-resonance “chemical presaturation” RF pulses to minimize the fat (or the silicone) signal in the image (right image of Fig. 15-36C). STIR (see Chapter 14) parameters can be selected to eliminate the signals due to fat at the “bounce point.” Additionally, swapping the phase and frequency gradient directions or changing the polarity of the frequency encode gradient can displace chemical shift artifacts from a specific image region. Identification of fat in a specific anatomic region is easily discerned from the chemical shift artifact displacement caused by changes in FEG/PEG gradient directions.

Ringing Artifacts

Ringing artifact (also known as the Gibbs phenomenon) occurs near sharp boundaries and high-contrast transitions in the image, and appears as multiple, regularly spaced parallel bands of alternating bright and dark signal that slowly fades with distance. The cause is the insufficient sampling of high frequencies inherent at sharp discontinuities in the signal. Images of objects can be reconstructed from a summation of sinusoidal waveforms of specific amplitudes and frequencies, as shown in Fig. 15-37A for a simple rectangular object. In the figure, the summation of frequency harmonics, each with a particular amplitude and phase, approximates the distribution of the object, but initially does very poorly, particularly at the sharp edges. As the number of higher frequency harmonics increase, a better estimate is achieved, although an infinite number of frequencies are theoretically necessary to reconstruct the sharp edge perfectly. In the MR acquisition, the number of frequency samples is determined by the number of pixels (frequency, k_x , or phase, k_y , increments) across the k-space matrix. For 256 pixels, 128 discrete frequencies are depicted, and for 128 pixels, 64 discrete frequencies are specified (the k-space matrix is symmetric in quadrants and duplicated about its center). A lack of high-frequency signals causes the “ringing” at sharp transitions described as a diminishing hyper- and hypointense signal oscillation from the transition. Ringing artifacts are thus more likely for smaller digital matrix sizes (Fig. 15-37C, 256 versus 128 matrix). Ringing artifact commonly occurs at skull/brain interfaces, where there is a large transition in signal amplitude.

Wraparound Artifacts

The wraparound artifact is a result of the mismapping of anatomy that lies outside of the FOV but within the slice volume. The anatomy is usually displaced to the opposite side of the image. It is caused by nonlinear gradients or by undersampling of the frequencies contained within the returned signal envelope. For the latter, the sampling rate must be twice the maximal frequency that occurs in the object (the

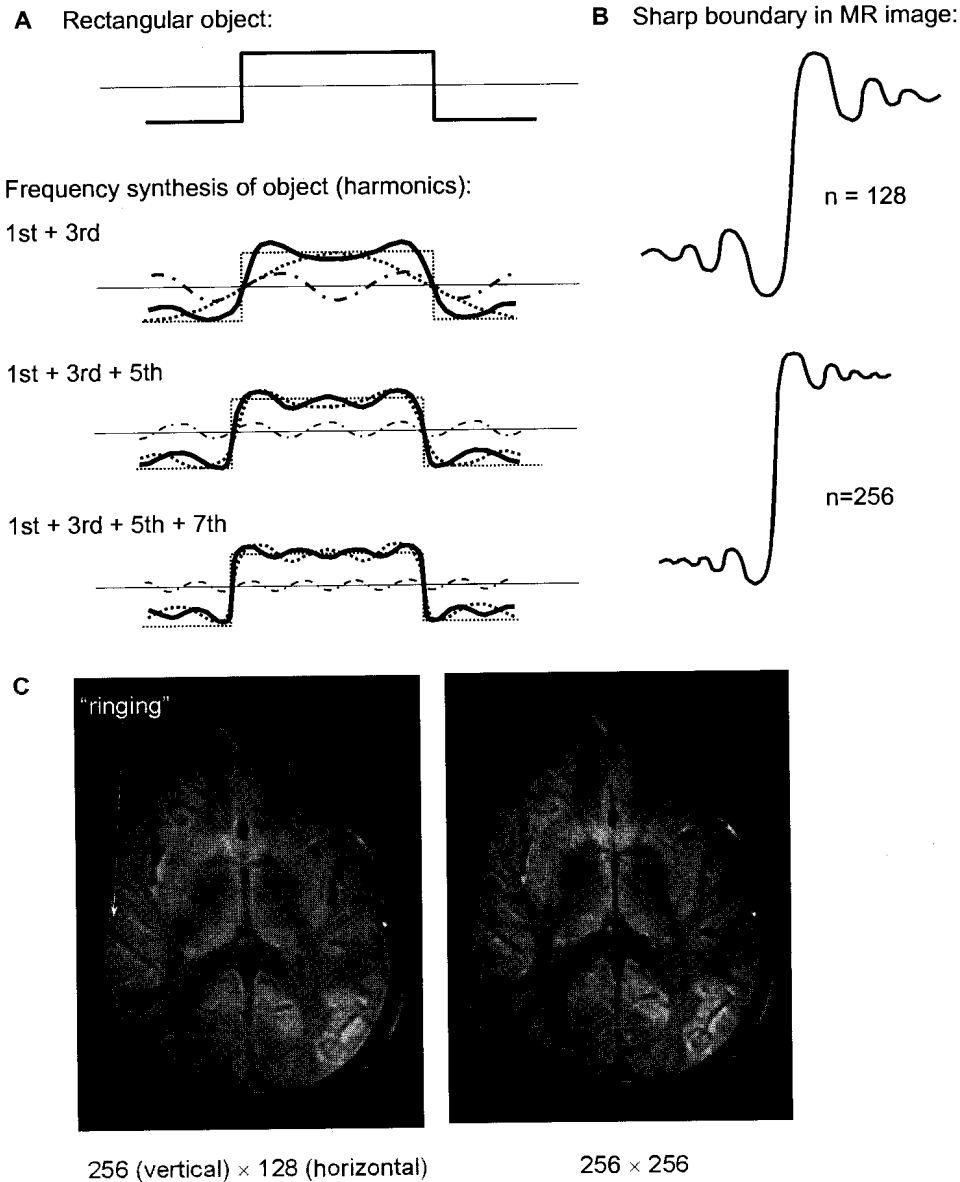


FIGURE 15-37. Ringing artifacts are manifested at sharp boundary transitions as a series of diminishing hyper- and hypointensities from the boundary, caused by a lack of sufficient sampling. **A:** The synthesis of a spatial object occurs by the summation of frequency harmonics in the MR image. As higher frequency harmonics are included, the summed result more faithfully represents the object shape. **B:** The number of frequencies encoded in the MR image is estimated with 256 samples better than with 128 samples; the amount of ringing is reduced with a larger number of samples. **C:** Example of ringing artifacts caused by a sharp signal transition at the skull in a brain image for a 256 $\times$ 128 matrix (*left*) along the short (horizontal) axis, and the elimination of the artifact in a 256 $\times$ 256 matrix (*right*). The short axis is along the PEG. (Images courtesy of Dr. Michael Buonocore, MD, PhD, University of California–Davis.)

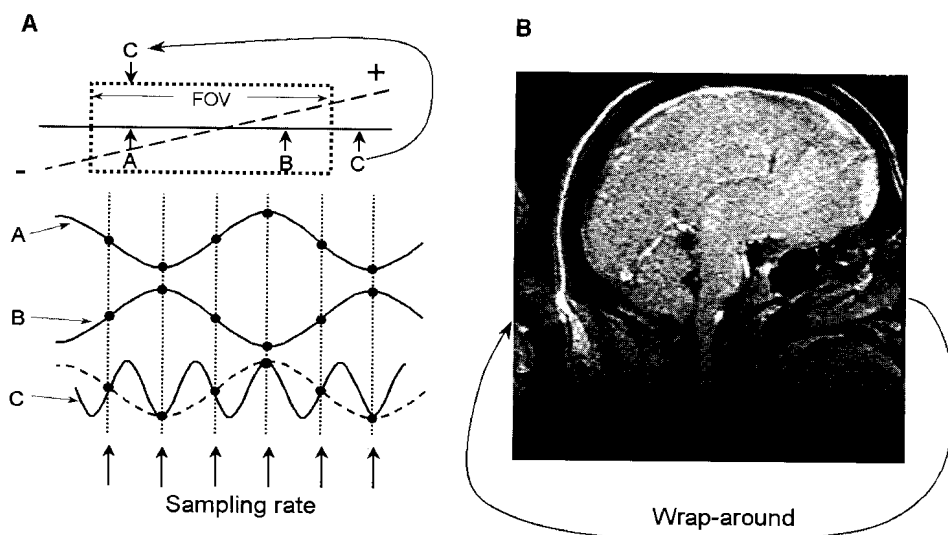


FIGURE 15-38. A: Wraparound artifacts are caused by aliasing. Shown is a fixed sampling rate and net precessional frequencies occurring at position A and position B within the FOV that have identical frequencies but different phase. If signal from position C is at twice the frequency of B but insufficiently sampled, the same frequency and phase will be assigned to C as that assigned to A, and therefore will appear at that location. **B:** A wraparound artifact example displaces anatomy from one side of the image (or outside of the FOV) to the other side.

Nyquist sampling limit). Otherwise, the Fourier transform cannot distinguish frequencies that are present in the data above the Nyquist frequency limit, and instead assigns a lower frequency value to them (Fig. 15-38A). Frequency signals will “wraparound” to the opposite side of the image, masquerading as low-frequency (aliased) signals (Fig. 15-38B).

In the frequency encode direction a low-pass filter can be applied to the acquired time domain signal to eliminate frequencies beyond the Nyquist frequency. In the phase encode direction aliasing artifacts can be reduced by increasing the number of phase encode steps (the trade-off is increased image time). Another approach is to move the region of anatomic interest to the center of the imaging volume to avoid the overlapping anatomy, which usually occurs at the periphery of the FOV. An “antialiasing” saturation pulse just outside of the FOV is yet another method of eliminating high-frequency signals that would otherwise be aliased into the lower-frequency spectrum.

Partial-Volume Artifacts

Partial-volume artifacts arise from the finite size of the voxel over which the signal is averaged. This results in a loss of detail and spatial resolution. Reduction of partial-volume artifacts is accomplished by using a smaller pixel size and/or a smaller slice thickness. With a smaller voxel, the SNR is reduced for a similar imaging time, resulting in a noisier signal with less low-contrast sensitivity. Of course, with a greater number of excitations (averages), the SNR can be maintained, at the cost of longer imaging time.

15.7 INSTRUMENTATION

Magnet

The magnet is the heart of the MR system. For any particular magnet type, performance criteria include field strength, temporal stability, and field homogeneity. These parameters are affected by the magnet design. Air core magnets are made of wire-wrapped cylinders of 1-m diameter and greater, where the magnetic field is produced by an electric current in the wires. The main magnetic field of air core magnets runs parallel to the long axis of the cylinder. Typically, the magnetic field is horizontal and runs along the cranial-caudal axis of the patient lying supine (Fig. 15-39A). Solid core magnets are constructed from permanent magnets, a wire wrapped iron core “electromagnet,” or a hybrid combination. In these solid core designs, the magnetic field runs between the poles of the magnet, most often in a vertical direction (Fig. 15-39B).

Resistive Magnet

Resistive electromagnets are constructed in either an air core or solid core configuration. Most resistive magnets are now of the solid core design and have a vertical magnetic field with contained fringe fields. These systems use continuous electric power to produce the magnetic field (must overcome the electric resistance of the coil wires), produce a significant amount of heat, and often require additional cooling subsystems. The magnetic field of resistive systems ranges from 0.1 T to about 0.3 T. An advantage of a

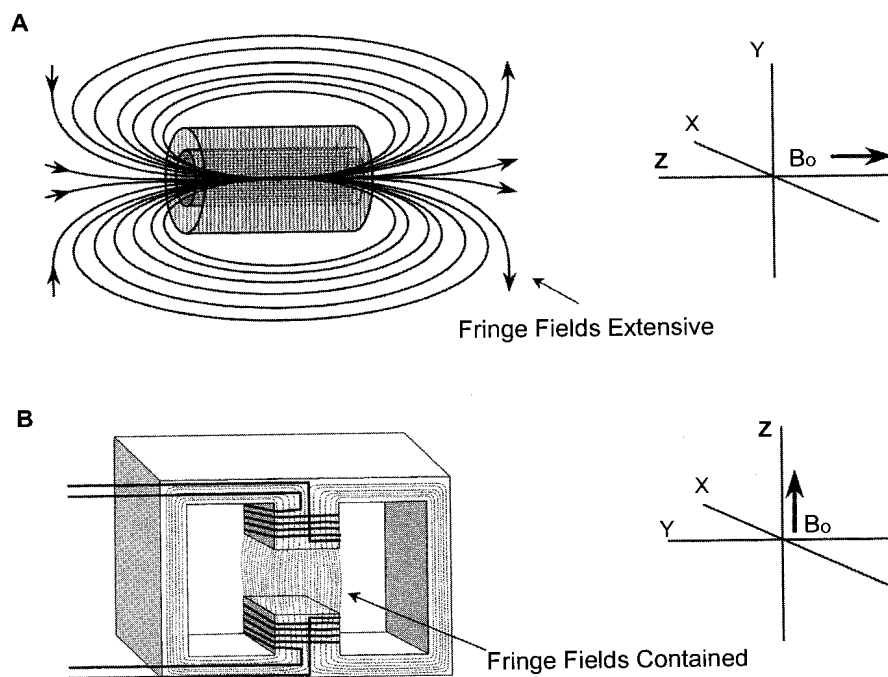


FIGURE 15-39. A: Air core magnets typically have a horizontal main field produced in the bore of the electric windings, with the z-axis (B_0) along the bore axis. Fringe fields for the air core systems are extensive and are increased for larger bore diameters and higher field strengths. **B:** The solid core magnet usually has a vertical field, produced between the metal poles of a permanent or wire-wrapped electromagnet. Fringe fields are confined with this design. The main field is parallel to the z-axis of the Cartesian coordinate system.

purely resistive system is the ability to turn off the magnetic field in an emergency. The disadvantages include high electricity costs and relatively poor uniformity/homogeneity of the field (e.g., 30 to 50 ppm in a 10 cm³ volume). Systems using an electromagnet solid core design overcome many of the disadvantages of the air core resistive magnet design; most significantly, the “open” design enables claustrophobic patients to tolerate the examination. Fringe fields of the main magnet are also better confined.

Superconductive Magnet

Superconductive magnets typically use an air core electromagnet configuration (there are some solid core superconductive magnets), and consist of a large cylinder of approximately 1 m in diameter and 2 to 3 m in depth, wrapped with a long, continuous strand of superconducting wire. Superconductivity is a characteristic of certain metals (e.g., niobium-titanium alloys) that exhibit no resistance to electric current when kept at extremely low temperatures, with liquid helium (boiling point of 4°K) required as the coolant. After initialization (“ramp up”) by an external electric source, current continues to flow as long as the wire temperature is kept at liquid helium temperatures. Higher temperatures cause the loss of superconductivity, and resistance heating “quenches” the magnetic field. Superconductive magnets achieve high field strengths (0.3- to 3.0-T clinical systems are common, and 4.0- to 7.0-T clinical large bore magnets are used in research). In addition, high field uniformity of 1 ppm in a 40-cm³ volume is typical. The several disadvantages of superconducting magnets include high initial capital and siting costs, cryogen costs (most magnets have refrigeration units to minimize liquid helium losses), difficulty in turning off the main magnetic field in an emergency, and extensive fringe fields. Uncontrolled quenching can cause explosive boiling of the liquid helium and severe damage to the magnet windings. Despite these disadvantages, the superconductive magnet is by far the most widely used magnet type for clinical imaging.

A cross section of the internal superconducting magnet components shows integral parts of the magnet system, including gradient, shim, and RF coils, as well as the cryogenic liquid containment vessels (Fig. 15-40).

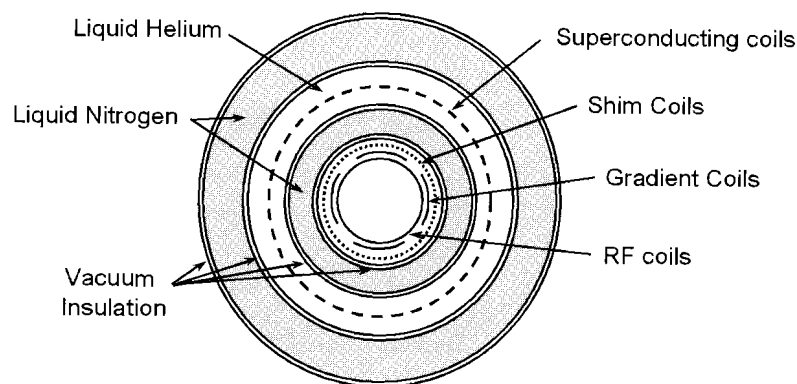


FIGURE 15-40. Cross-sectional view of the superconducting magnet shows the coils for the magnet, shims, RF, and gradients. Surrounding the superconductor coil assembly is the liquid helium bath and surrounding wrap of liquid nitrogen containers with vacuum insulation barriers. In modern systems with refrigeration, the liquid nitrogen containers are unnecessary.

Permanent Magnet

Permanent magnets rely on the ferromagnetic properties of iron, nickel, cobalt, and alloys. Early permanent magnets were extremely large, bulky and heavy, some weighing over 100 tons. Lighter alloy permanent magnets with improved peripheral electronic equipment are finding a niche in clinical MRI. Operational costs of these systems are the lowest of all MR systems (no major electric or cooling requirements). Field uniformity is typically less than a superconducting magnet with similar FOV. An inability to turn off the field in an emergency is a drawback.

Ancillary Equipment

Ancillary components are necessary for the proper functioning of the MR system. Shim coils are active or passive magnetic field devices that are used to adjust the main magnetic field and to improve the homogeneity in the sensitive central volume of the scanner. Gradient coils, as previously discussed, are wire conductors that produce a linear superimposed gradient magnetic field on the main magnetic field. The gradient coils are located inside the cylinder of the magnet, and are responsible for the banging noise one hears during imaging. This noise is caused by the flexing and torque experienced by the gradient coil from the rapidly changing magnetic fields when energized.

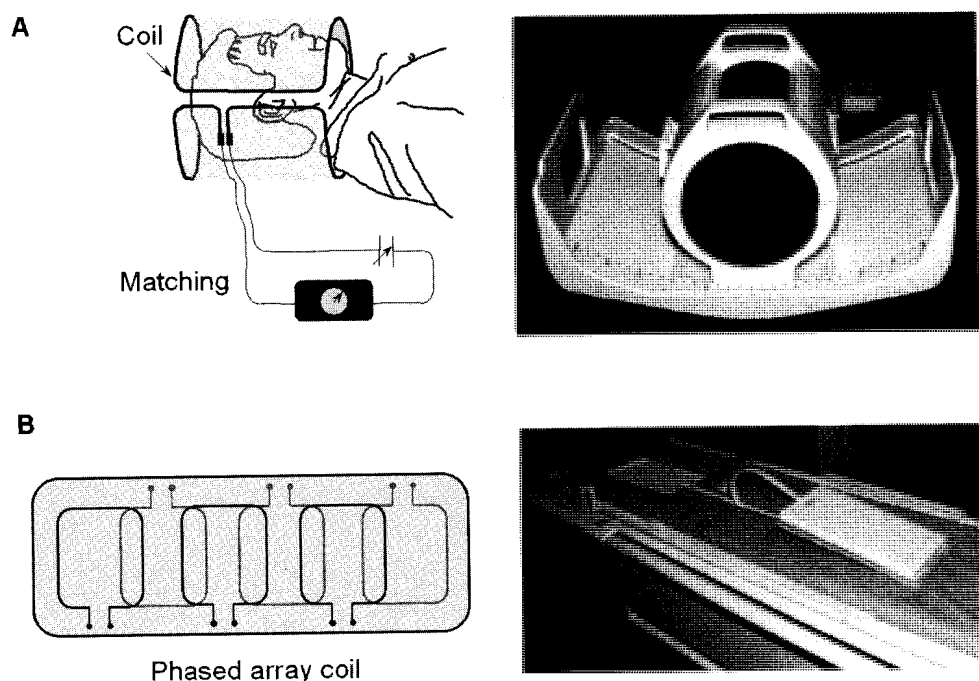


FIGURE 15-41. A: Proximity head coil design and photograph of clinical head coil. The matching circuitry provides tuning of the coil for improved sensitivity and response. Two or more tuning capacitors are required. **B:** A multiple phased array surface coil detector provides extended coverage and SNR. In this design, a subset of the coils are activated at any one time (indicated by darker lines); postprocessing is necessary to combine the images.

RF coils create the B_1 field that rotates the magnetization in a pulse sequence, and detect the transverse magnetization as it precesses in the x-y plane. RF transmitter and receiver body coils are located within the magnet bore. There are two types of RF coils: transmit and receive, and receive only. Often, transmit and receive functions are separated to handle the variety of imaging situations that arise, and to maximize the SNR for an imaging sequence. All coils must resonate and efficiently store energy at the Larmor frequency. This is determined by the inductance and capacitance properties of the coil. RF coils need to be tuned prior to each acquisition and matched to accommodate the different magnetic inductance of each patient. Proximity RF coils include bird-cage coils, the design of choice for brain imaging, the single-turn solenoid for imaging the extremities and the breasts, and the saddle coil. These coils are typically operated as both a transmitter and receiver of RF energy (Fig. 15-41A). Surface coils and phased-array coils are used in lieu of the body coil when high-resolution or smaller FOV areas are desired. As a receive-only coil, surface and phased array coils have a high SNR for adjacent tissues, although the field intensity drops off rapidly at the periphery. The phased array coil is made of several overlapping loops, which extend the imaging FOV in one direction (Fig. 15-41B). Quadrature detector coils are sensitive to changing magnetic fields along two axes. This detector manages two simultaneous channels known as the real (records MR information in phase with a reference signal) and the imaginary (records MR information 90 degree out of phase with the reference signal) channels. This detector increases the SNR by a factor of $\sqrt{2}$. If imbalances in the offset or gain of these detectors occur, then artifacts will be manifested, such as a "center point" artifact.

The control interfaces, RF source, detector, and amplifier, analog to digital converter (digitizer), pulse programmer, computer system, gradient power supplies, and image display are crucial components of the MR system. They integrate and synchronize the tasks necessary to produce the MR image (Fig. 15-42).

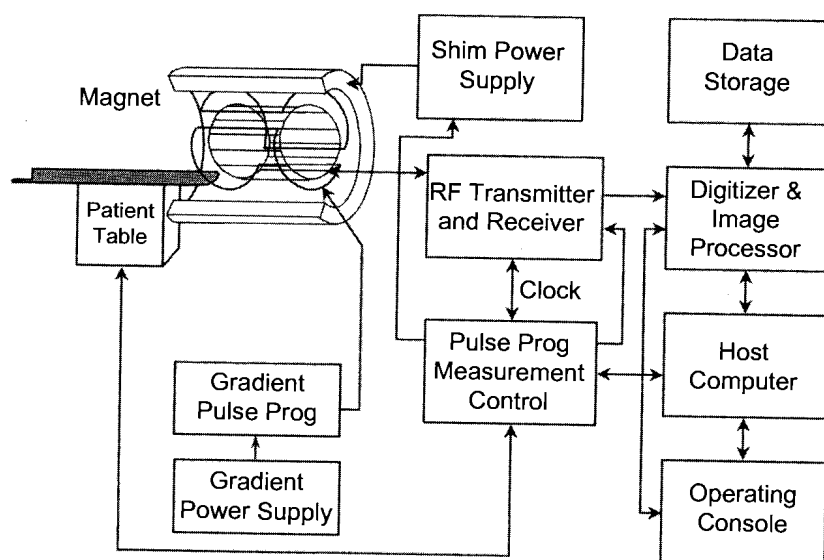


FIGURE 15-42. A block diagram of a typical MRI system shows the components and interfaces between subsystems. Not shown is the Faraday cage that surrounds the magnet assembly to eliminate environmental RF noise.

Magnet Siting

Superconductive magnets produce extensive magnetic fringe fields, and create potentially hazardous conditions in adjacent areas. In addition, extremely small signal amplitudes generated by the protons in the body during an imaging procedure have a frequency common to commercial FM broadcasts. Thus, two requirements must be considered for MR system siting: protect the local environment from the magnet system, and protect the magnet system from the local environment.

Fringe fields from a high field strength magnet can extend quite far—roughly equal to αB_0 , where α is a constant dependent on the magnet bore size and magnet configuration. Fringe fields can potentially cause a disruption of electronic signals and sensitive electronic devices. An unshielded 1.5-T magnet has a 1-mT (10 G) fringe field at a distance of approximately 9.3 m, a 0.5-mT (5 G) field at 11.5 m, and a 0.3-mT (3 G) field at 14.1 m from the center of the magnet. Magnetic shielding is one way to reduce fringe field interactions in adjacent areas. Passive (e.g., thick metal walls close to the magnet) and active (e.g., electromagnet systems strategically placed in the magnet housing) magnetic shielding systems permit a significant reduction in the extent of the fringe fields for high field strength, air core magnets (Fig. 15-43A). Patients with pacemakers or ferromagnetic aneurysm clips must avoid fringe fields above 0.5 mT. Magnetically sensitive equipment such as image intensifiers, gamma cameras, and color TVs are severely impacted by fringe fields of less than 0.3 mT, as electromagnetic focusing in these devices is disrupted.

Administrative control for magnetic fringe fields is 0.5 mT (5 G), requiring controlled access to areas that exceed this level (Fig. 15-43B). Magnetic fields below 0.5 mT are considered safe for the patient population. Areas above 1.0 mT require controlled and restricted access with warning signs. Disruption of the fringe fields can reduce the homogeneity of the active imaging volume. Any large metallic object (e.g., elevator, automobile, etc.) traveling through the fringe field can produce such an effect.

Environmental RF noise must be reduced to protect the extremely sensitive receiver within the magnet from interfering signals. One approach for stray RF signal protection is the construction of a Faraday cage, an internal enclosure consisting of RF attenuating copper sheet and/or copper wire mesh. The room containing the MRI system is typically lined with copper sheet (walls) and mesh (windows). This is a costly construction item but provides effective protection from stray RF noise. Another method is the integration of a “waveguide” tunnel into the scanner itself. A half-cylinder-shaped cover is pulled over the patient beyond the opening of the imaging bore. Attenuation of environmental RF noise is significant, although typically not as complete as the Faraday cage.

Quality Control

Like any imaging system, the MRI scanner is only as good as the weakest link in the imaging chain. The components of the MR system that must be periodically checked include the magnetic field strength, magnetic field homogeneity, system field shimming, gradient linearity, system RF tuning, receiver coil optimization, environmental noise sources, power supplies, peripheral equipment, and control systems among others. A quality control (QC) program should be designed to assess the basic day-to-day functionality of the scanner. A set of QC tests is listed in Table

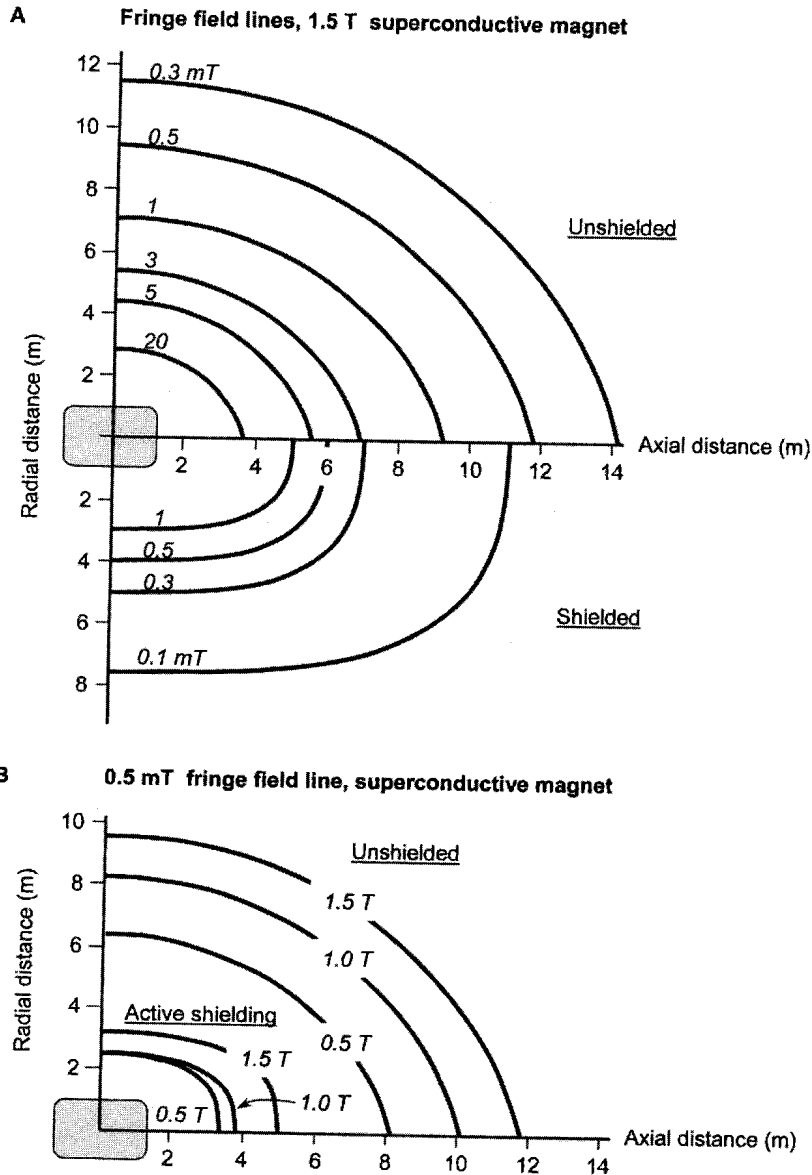


FIGURE 15-43. A: Magnetic fringe fields (mT) from a 1.5-T magnet indicate the extent of the field strength axially and radially for an unshielded and passively shielded imaging system. The fringe fields extend symmetrically about the magnet. **B:** The 0.5-mT fringe field line for administrative control is indicated for 0.5-, 1.0-, and 1.5-T magnets with no shielding and active shielding.

TABLE 15-2. RECOMMENDED QUALITY CONTROL TESTS FOR MAGNETIC RESONANCE IMAGING SYSTEMS

High-contrast spatial resolution
Slice thickness accuracy
Radiofrequency bandwidth tuning
Geometric accuracy and spatial uniformity
Signal uniformity
Low-contrast resolution (sensitivity)
Image artifact evaluation
Preventive maintenance logging and documentation
Review of system log book and operations

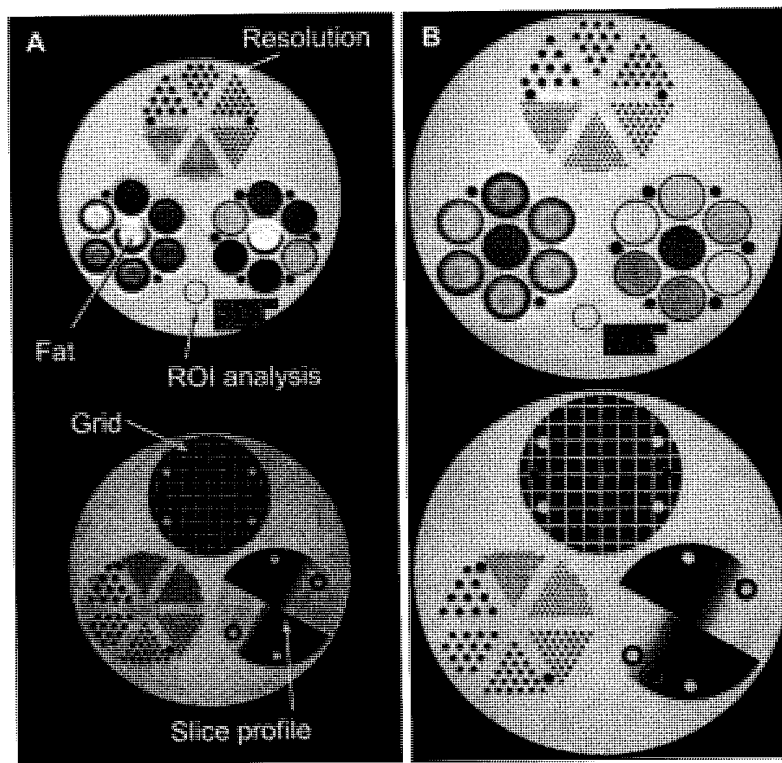


FIGURE 15-44. Images of an MRI quality control (QC) phantom indicate performance of the MRI system for spatial resolution, slice thickness, SNR, distortion, distance accuracy, and chemical shift artifacts among other measurements. SNR is calculated from a region of interest in the images as average number/standard deviation. Various magnetic susceptibility agents are placed in the vials to produce known T1 and T2 values; of note is the center vial filled with oil, and the easily recognized chemical shift artifacts. **A:** SPGR (TR = 10.2 msec, TE = 1.8 msec, slice = 2 mm, 256 × 128 image, SNR = 55.7). **B:** Three-dimensional gradient recalled echo (3D GRE) (TR = 31.0 msec, TE = 10.0, slice = 4 mm, 512 × 512 image, SNR = 64.0).

15-2. Qualitative and quantitative measurements of system performance should be obtained on a periodic basis, with a test frequency dependent on the likelihood of detecting a change in the baseline values outside of normal operating limits.

The American College of Radiology has an MRI accreditation program that specifies requirements for system operation, quality control, and the training requirements of technologists, radiologists, and physicists involved in scanner operation. The accreditation process evaluates the qualifications of personnel, equipment performance, effectiveness of QC procedures, and quality of clinical images—factors that are consistent with the maintenance of a state-of-the-art facility.

MRI phantoms are composed of materials that produce MR signals with carefully defined relaxation times. Some materials are aqueous paramagnetic solutions; pure gelatin, agar, silicone, or agarose; and organic doped gels, paramagnetic doped gels, and others. Water is most frequently used, but it is necessary to adjust the T1 and T2 relaxation times of (doped) water so that images can be acquired using pulse sequence timing for patients (e.g., this is achieved by adding nickel, aqueous oxygen, aqueous manganese, or gadolinium). A basic MR phantom can be used to test resolution (contrast and spatial), slice thickness, and RF uniformity/homogeneity. The performance metrics measured include spatial properties such as in-plane resolution, slice thickness, linearity, and SNR as a function of position (Fig. 15-44). Some phantoms have standards with known T1, T2, and spin density values to evaluate the quantitative accuracy of the scanner, and to determine the ability to achieve an expected contrast level for a given pulse sequence. Homogeneity phantoms determine the spatial uniformity of transmit and receive RF magnetic fields. Ideal performance is a spatially uniform excitation of the spins and a spatially uniform sensitivity across the imaged object.

15.8 SAFETY AND BIOEFFECTS

Although ionizing radiation is not used with MRI, there are important safety considerations. These include the presence of strong magnetic fields, RF energy, time-varying magnetic gradient fields, cryogenic liquids, a confined imaging device (claustrophobia), and noisy operation (gradient coil activation and deactivation, creating acoustic noise). Patients with implants, prostheses, aneurysm clips, pacemakers, heart valves, etc., should be aware of considerable torque when placed in the magnetic field, which could cause serious adverse effects. Even nonmetallic implant materials can lead to significant heating under rapidly changing gradient fields. Consideration of the distortions on the acquired images and possibility of misdiagnosis are also a concern. Ferromagnetic materials inadvertently brought into the imaging room (e.g., an IV pole) are attracted to the magnetic field and can become a deadly projectile to the occupant within the bore of the magnet.

The long-term biologic effects of high magnetic field strengths are not well known. At lower magnetic field strengths, there have not been any reports of deleterious biologic effects, either acute or chronic. With very high field strength magnets (e.g., 4 T or higher), there has been anecdotal mention of dizziness and disorientation of personnel and patients as they move through the field. The most common bioeffect of MR systems is tissue heating caused by RF energy deposition and/or by rapid switching of high strength gradients.

TABLE 15-3. MAGNETIC RESONANCE IMAGING (MRI) SAFETY GUIDELINES (U.S. FOOD AND DRUG ADMINISTRATION)

Issue	Parameter	Variables	Specified value
Static magnetic field	Magnetic field (B_0) Inadvertent exposure	Maximum strength	2.0 T ^a
		Maximum	0.0005 T
Changing magnetic field (dB/dt)	Axial gradients	$\tau > 120 \mu\text{sec}$	$< 20 \text{ T/sec}$
		$12 \mu\text{sec} < \tau < 120 \mu\text{sec}$	$< 2400/\tau (\mu\text{s}) \text{ T/sec}$
		$\tau < 12 \mu\text{sec}$	$< 200 \text{ T/sec}$
Radiofrequency power deposition	Transverse gradients System		$< 3 \times \text{axial gradients}$
			$< 6 \text{ T/sec}$
	Temperature	Change	$< 1^\circ\text{C}$ local
		Maximum head	$< 38^\circ\text{C}$
		Maximum trunk	$< 39^\circ\text{C}$
	Specific absorption rate (SAR)	Maximum extremities	$< 40^\circ\text{C}$
		Whole body	$< 0.4 \text{ W/kg}$
		Any 1 g of tissue	$< 8.0 \text{ W/kg}$
Acoustic noise levels		Head (averaged)	$< 3.2 \text{ W/kg}$
		Peak pressure	200 pascals
		Average pressure	105 dBA

^aIn some clinical applications, higher field strengths (e.g., 3.0 T) are allowed.
 τ , rise time of gradients.

MRI is considered extremely safe when used within the regulatory guidelines (Table 15-3). Serious bioeffects are demonstrated with static and varying magnetic fields at strengths significantly higher ($10\times$ to $20\times$) than those used for typical diagnostic imaging.

Static Magnetic Fields

Very high field strength static magnetic fields have been shown to increase membrane permeability in the laboratory. With systems in excess of 20 T, enzyme kinetic changes have been documented, and altered biopotentials have been measured. These effects have not been demonstrated in magnetic fields below 10 T.

Varying Magnetic Field Effects

The time-varying magnetic fields encountered in the MRI system are due to the gradient switching used for localization of the protons. Rapid quenching of the B_0 field can also cause a time-varying magnetic field. Magnetic fields that vary their strength with time are generally of greater concern than static fields because oscillating magnetic fields can induce electrical current flow in conductors. (See Chapter 5 on transformers for a more detailed explanation.) The maximum allowed changing gradient fields depend on the rise times of the gradients, as listed in Table 15-3. At extremely high levels of magnetic field variation, effects such as visual phosphenes (the sensation of flashes of light being seen) can result because of induced currents in the nerves or tissues. Other consequences such as bone healing and cardiac fibrillation have been suggested in the literature.

Magnetic Field, RF Exposure, and Noise Limits

RF exposure causes heating of tissues. There are obvious effects of overheating, and therefore a power deposition limit is imposed by governmental regulations for various aspects of MRI and MRS operation. Table 15-3 lists some of the categories and the maximum values permitted for clinical use. The rationale for imposing limit on static and varying magnetic fields is based on the ability of the resting body to dissipate heat buildup caused by the deposition and absorption of thermal energy.

SUGGESTED READING

- NessAiver M. *All you really need to know about MRI physics*. Baltimore, MD: Simply Physics, 1997.
- Price RR. The AAPM/RSNA physics tutorial for residents: MR imaging safety considerations. *RadioGraphics* 1999;19:1641–1651.
- Saloner D. The AAPM/RSNA physics tutorial for residents. An introduction to MR angiography. *RadioGraphics* 1995;15:453–465.
- Shellock F, Kanal E. *Magnetic resonance imaging bioeffects, safety, and patient management*, 2nd ed. New York: Lippincott-Raven, 1996.
- Smith HJ, Ranallo FN. *A non-mathematical approach to basic MRI*. Madison, WI: Medical Physics, 1989.

ULTRASOUND

Ultrasound is the term that describes sound waves of frequencies exceeding the range of human hearing and their propagation in a medium. Medical diagnostic ultrasound is a modality that uses ultrasound energy and the acoustic properties of the body to produce an image from stationary and moving tissues. Ultrasound imaging uses a “pulse echo” technique to synthesize a gray-scale tomographic image of tissues based on the mechanical interaction of short pulses of high-frequency sound waves and their returning echoes. Generation of the sound pulses and detection of the echoes is accomplished with a transducer, which also directs the ultrasound pulse along a linear path through the patient. Along a given beam path, the depth of an echo-producing structure is determined from the time between the pulse emission and the echo return, and the amplitude of the echo is encoded as a gray-scale value (Fig. 16-1). In addition to two-dimensional (2D) tomographic imaging, ultrasound provides anatomic distance and volume measurements, motion studies, blood velocity measurements, and three-dimensional (3D) imaging.

Historically, medical uses of ultrasound came about shortly after the close of World War II, derived from underwater sonar research. Initial clinical applications monitored changes in the propagation of pulses through the brain to detect intracerebral hematoma and brain tumors based on the displacement of the midline. Ultrasound rapidly progressed through the 1960s from simple “A-mode” scans to “B-mode” applications and compound “B-scan” images using analog electronics. Advances in equipment design, data acquisition techniques, and data processing capabilities have led to electronic transducer arrays, digital electronics, and real-time image display. This progress is changing the scope of ultrasound and its applications in diagnostic radiology and other areas of medicine. High-resolution, real-time imaging, harmonic imaging, 3D data acquisition, and power Doppler are a few of the innovations introduced into clinical practice. Contrast agents for better delineation of the anatomy, measurement of tissue perfusion, precise drug delivery mechanisms, and determination of elastic properties of the tissues are topics of current research.

This chapter describes the characteristics, properties, and production of ultrasound; the modes of ultrasound interaction, instrumentation, and image acquisition; signal processing; image display; Doppler flow measurement; quality control; and ultrasound bioeffects.

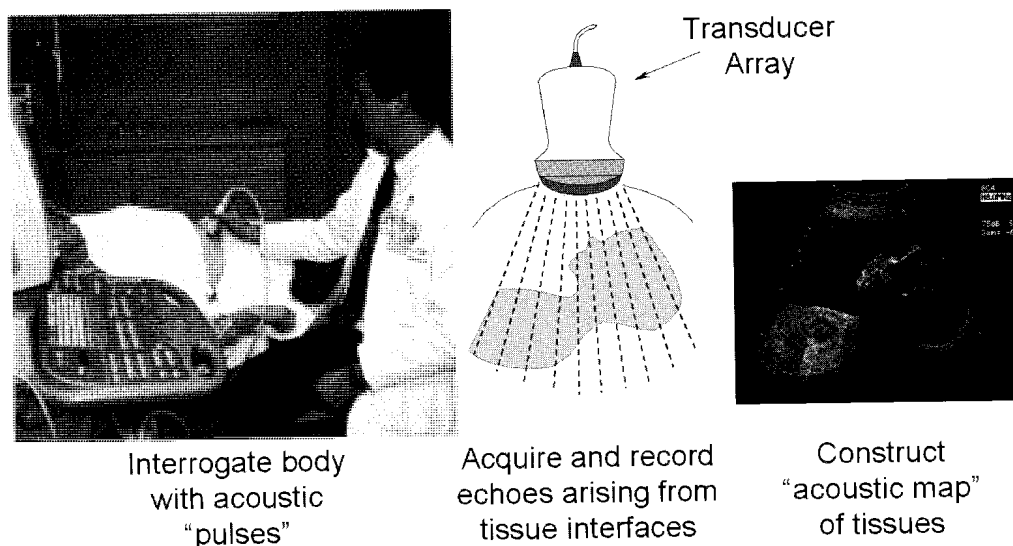


FIGURE 16-1. A major use of ultrasound is the acquisition and display of the acoustic properties of tissues. A transducer array (transmitter and receiver of ultrasound pulses) directs sound waves into the patient, receives the returning echoes, and converts the echo amplitudes into a two-dimensional tomographic image. Exams requiring a high level of safety such as obstetrics are increasing the use of ultrasound in diagnostic radiology.

16.1 CHARACTERISTICS OF SOUND

Propagation of Sound

Sound is mechanical energy that propagates through a continuous, elastic medium by the compression and rarefaction of “particles” that compose it. A simple model of an elastic medium is that of a spring (Fig. 16-2). Compression is caused by a mechanical deformation induced by an external force (such as a plane piston), with a resultant increase in the pressure of the medium. Rarefaction occurs following the compression event; as the backward motion of the piston reverses the force, the compressed particles transfer their energy to adjacent particles, with a subsequent reduction in the local pressure amplitude. The section of spring in contact with the piston then becomes stretched (a rarefaction) while the compression continues to travel forward. With continued back-and-forth motion of the piston, a series of compressions and rarefactions propagate through the continuous medium (the spring). Any change with respect to position in the elasticity of the spring disturbs the wave train and causes some or all of the energy to be reflected. While the medium itself is necessary for mechanical energy transfer (i.e., sound propagation), the constituent “particles” of the medium act only to transfer mechanical energy; these particles experience only very small back-and-forth displacements. Energy propagation occurs as a wave front in the direction of energy travel, known as a longitudinal wave.

Sound energy can also be produced in short bursts, such that a small pulse travels by itself through the medium. Echoes, reflections of the incident energy pulse, occur due to differences in the elastic properties of the medium. An acoustic image is formed from numerous pulses of ultrasound reflected from tissue interfaces back to the receiver.

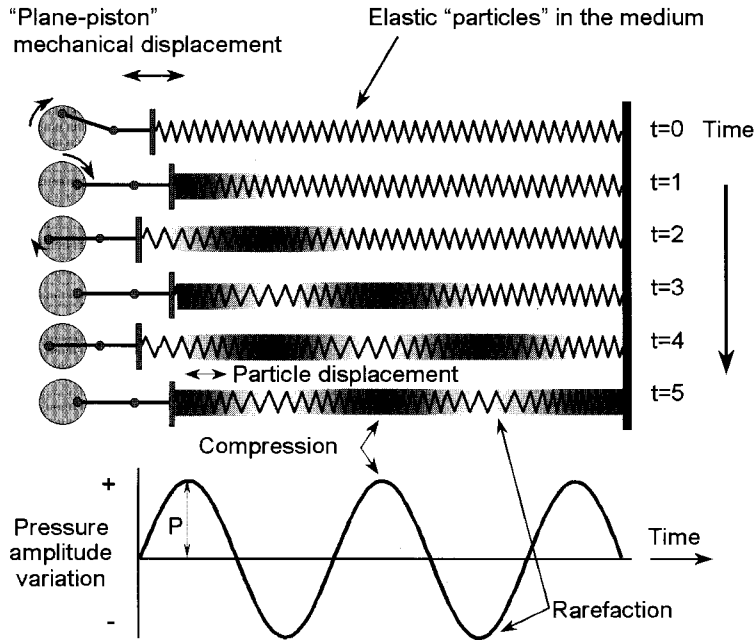


FIGURE 16-2. Ultrasound energy is generated by a mechanical displacement in compressible medium, which is modeled as an elastic spring. Energy propagation is shown as a function of time, resulting in areas of compression and rarefaction with corresponding variations in positive and negative pressure amplitude, P (lower diagram, a snapshot of the ultrasound wave at $t=5$ above). Particles in the medium move back and forth in the direction of the wave front.

Wavelength, Frequency, and Speed

The wavelength (λ) of the ultrasound is the distance (usually expressed in millimeters or micrometers) between compressions or rarefactions, or between any two points that repeat on the sinusoidal wave of pressure amplitude (Fig. 16-2). The frequency (f) is the number of times the wave oscillates through a cycle each second (sec). Sound waves with frequencies less than 15 cycles/sec (Hz) are called infrasound, and the range between 15 Hz and 20 kHz comprises the audible acoustic spectrum. Ultrasound represents the frequency range above 20 kHz. Medical ultrasound uses frequencies in the range of 2 to 10 MHz, with specialized ultrasound applications up to 50 MHz. The period is the time duration of one wave cycle, and is equal to $1/f$, where f is expressed in cycles/sec. The speed of sound is the distance traveled by the wave per unit time and is equal to the wavelength divided by the period. Since period and frequency are inversely related, the relationship between speed, wavelength, and frequency for sound waves is

$$c = \lambda f$$

where c (m/sec) is the speed of sound of ultrasound in the medium, λ (m) is the wavelength, and f (cycles/sec) is the frequency.

The speed of sound is dependent on the propagation medium and varies widely in different materials. The wave speed is determined by the ratio of the bulk mod-

ulus (B) (a measure of the stiffness of a medium and its resistance to being compressed), and the density (ρ) of the medium:

$$c = \sqrt{\frac{B}{\rho}}$$

SI units are $\text{kg}/(\text{m}\cdot\text{sec}^2)$, kg/m^3 , and m/sec for B, ρ , and c , respectively. A highly compressible medium, such as air, has a low speed of sound, while a less compressible medium, such as bone, has a higher speed of sound. A less dense medium has a higher speed of sound than a denser medium (e.g., dry air vs. humid air). The speeds of sound in materials encountered in medical ultrasound are listed in Table 16-1. Of major importance are the speed of sound in air (330 m/sec), the average speed for soft tissue (1,540 m/sec), and fatty tissue (1,450 m/sec). The difference in the speed of sound at tissue boundaries is a fundamental cause of contrast in an ultrasound image. Medical ultrasound machines assume a speed of sound of 1,540 m/sec . The speed of sound in soft tissue can be expressed in other units such as 154,000 cm/sec and 1.54 $\text{mm}/\mu\text{sec}$, and these values are often helpful in simplifying calculations.

The ultrasound frequency is unaffected by changes in sound speed as the acoustic beam propagates through various media. Thus, the ultrasound wavelength is dependent on the medium. Examples of the wavelength as a function of frequency and propagation medium are given below.

Example: A 2-MHz beam has a wavelength in soft tissue of

$$\lambda = \frac{c}{f} = \frac{1,540 \text{ m/sec}}{2 \times 10^6/\text{sec}} = 770 \times 10^{-6} = 7.7 \times 10^{-4} \times 1,000 \frac{\text{mm}}{\text{m}} = 0.77 \text{ mm}$$

A 10-MHz ultrasound beam has a corresponding wavelength in soft tissue of

$$= \frac{1,540 \text{ m/sec}}{10 \times 10^6/\text{sec}} = 154 \times 10^{-6} = 1.54 \times 10^{-4} \times 1,000 \frac{\text{mm}}{\text{m}} \approx 0.15 \text{ mm}$$

So, higher frequency sound has shorter wavelength (Fig. 16-3).

TABLE 16-1. DENSITY AND SPEED OF SOUND IN TISSUES AND MATERIALS FOR MEDICAL ULTRASOUND

Material	Density (kg/m^3)	c (m/s)	c (mm/ μs)
Air	1.2	330	0.33
Lung	300	600	0.60
Fat	924	1,450	1.45
Water	1,000	1,480	1.48
Soft tissue	1,050	1,540	1.54
Kidney	1,041	1,565	1.57
Blood	1,058	1,560	1.56
Liver	1,061	1,555	1.55
Muscle	1,068	1,600	1.60
Skull bone	1,912	4,080	4.08
PZT	7,500	4,000	4.00

PZT, lead-zirconate-titanate.

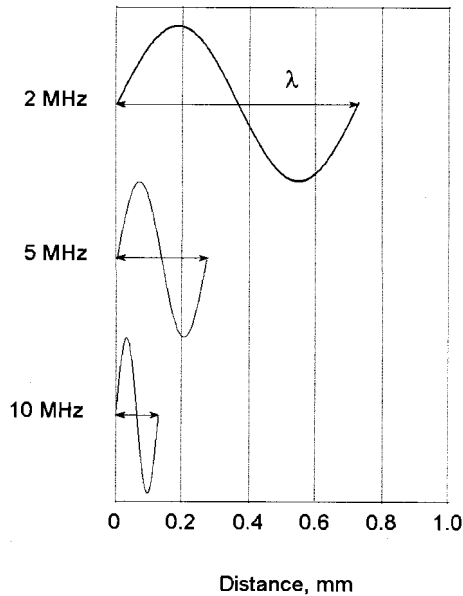


FIGURE 16-3. Ultrasound wavelength is determined by the frequency and the speed of sound in the propagation medium. Wavelengths in soft tissue are shown for 2-MHz, 5-MHz, and 10-MHz ultrasound.

Example: A 5-MHz beam travels from soft tissue into fat. Calculate the wavelength in each medium, and determine the percent wavelength change.

In soft tissue,

$$\lambda = \frac{c}{f} = \frac{1,540 \text{ m/sec}}{5 \times 10^6/\text{sec}} = 3.08 \times 10^{-6} \approx 0.31 \text{ mm}$$

In fat,

$$= \frac{1,450 \text{ m/sec}}{5 \times 10^6/\text{sec}} = 2.9 \times 10^{-6} \approx 0.29 \text{ mm}$$

A decrease in wavelength of 5.8% occurs in going from soft tissue into fat, due to the differences in the speed of sound.

The wavelength in mm in soft tissue can be calculated from the frequency specified in MHz using the approximate speed of sound in soft tissue ($c = 1540 \text{ m/sec} = 1.54 \text{ mm}/\mu\text{sec}$):

$$\lambda \text{ (mm, soft tissue)} = \frac{c}{f} = \frac{1,540 \text{ mm}/\mu\text{sec}}{f(\text{MHz})} = \frac{1,540 \text{ mm}/10^{-6} \text{ sec}}{f(10^6/\text{sec})} = \frac{1.54 \text{ mm}}{f(\text{MHz})}$$

A change in speed at an interface between two media causes a change in wavelength.

The resolution of the ultrasound image and the attenuation of the ultrasound beam energy depend on the wavelength and frequency. Ultrasound wavelength determines the spatial resolution achievable along the direction of the beam. A high-frequency ultrasound beam (small wavelength) provides superior resolution and image detail than a low-frequency beam. However, the depth of beam penetration is reduced at higher frequency. Lower frequency ultrasound has longer wavelength and less resolution, but a greater penetration depth. Ultrasound frequencies selected for imaging

are determined by the imaging application. For thick body parts (e.g., abdominal imaging), a lower frequency ultrasound wave is used (3.5 to 5 MHz) to image structures at significant depths, whereas for small body parts or organs close to the skin surface (e.g., thyroid, breast), a higher frequency is employed (7.5 to 10 MHz). Most medical imaging applications use frequencies in the range of 2 to 10 MHz.

Modern ultrasound equipment consists of multiple sound transmitters that create sound beams independent of each other. Interaction of two or more separate ultrasound beams in a medium results in constructive and/or destructive wave interference (Fig. 16-4). Constructive wave interference results in an increase in the amplitude of the beam, while destructive wave interference results in a loss of amplitude. The amount of constructive or destructive interference depends on several factors, but the most important are the phase (position of the periodic wave with respect to a reference point) and amplitude of the interacting beams. When the beams are exactly in phase and at the same frequency, the result is the constructive addition of the amplitudes (Fig. 16-4A). For equal frequency and a 180-degree phase difference, the result will be the destructive subtraction of the resultant beam amplitude (Fig. 16-4B). With phase and frequency differences, the results of the beam interaction can generate a complex interference pattern (Fig. 16-4C). The constructive and destructive interference phenomena are very important in shaping and steering the ultrasound beam.

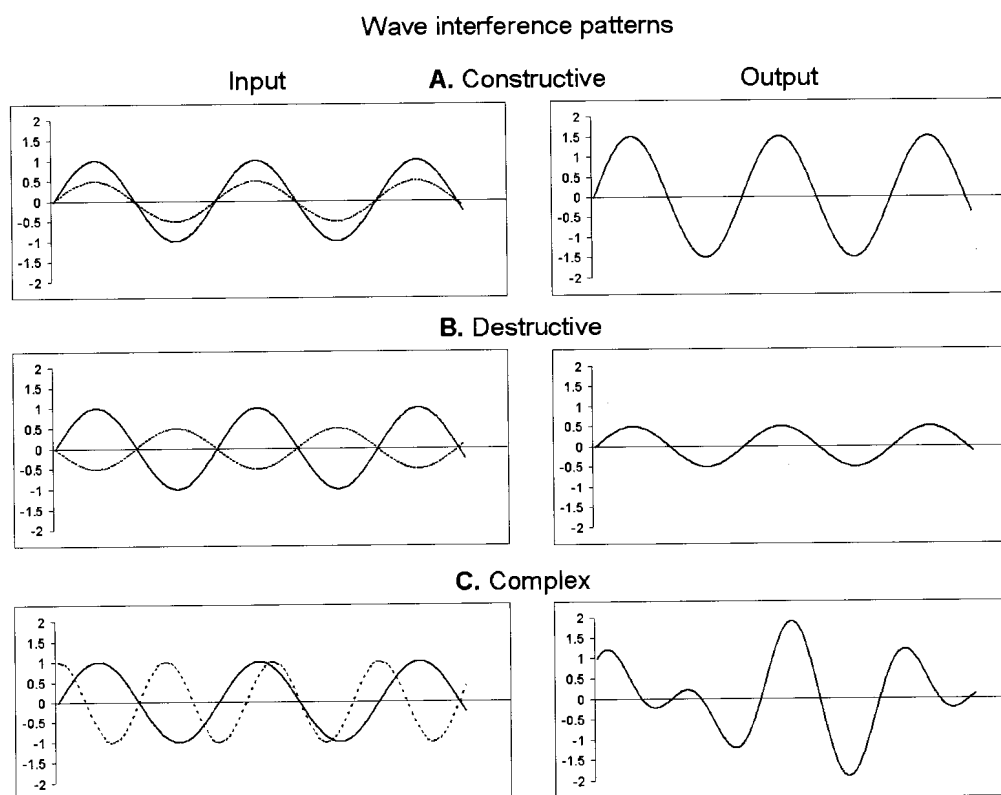


FIGURE 16-4. Interaction of waves in a propagation medium, plotted as amplitude versus time. **A:** Constructive interference occurs with two ultrasound waves of the same frequency and phase, resulting in a higher amplitude output wave. **B:** Destructive interference occurs with the waves 180 degrees out of phase, resulting in a lower amplitude output wave. **C:** Complex interference occurs when waves of different frequency interact, resulting in areas of higher and lower amplitude.

Pressure, Intensity, and the dB Scale

Sound energy causes particle displacements and variations in local pressure in the propagation medium. The pressure variations are most often described as pressure amplitude (P). Pressure amplitude is defined as the peak maximum or peak minimum value from the average pressure on the medium in the absence of a sound wave. In the case of a symmetrical waveform, the positive and negative pressure amplitudes are equal (lower portion of Fig. 16-2); however, in most diagnostic ultrasound applications, the compressional amplitude significantly exceeds the rarefactional amplitude. The SI unit of pressure is the pascal (Pa), defined as one newton per square meter (N/m^2). The average atmospheric pressure on earth at sea level of 14.7 pounds per square inch is approximately equal to 100,000 Pa. Diagnostic ultrasound beams typically deliver peak pressure levels that exceed ten times the earth's atmospheric pressure, or about 1 MPa (megapascal).

Intensity, I , is the amount of power (energy per unit time) per unit area and is proportional to the square of the pressure amplitude:

$$I \propto P^2$$

A doubling of the pressure amplitude quadruples the intensity. Medical diagnostic ultrasound intensity levels are described in units of milliwatts/ cm^2 —the amount of energy per unit time per unit area. The absolute intensity level depends on the method of ultrasound production (e.g., pulsed or continuous; a discussion of these parameters can be found in section 16.11 of this chapter). Relative intensity and pressure levels are described with a unit termed the decibel (dB). The relative intensity in dB is calculated as

$$\text{Relative intensity (dB)} = 10 \log \left(\frac{I_2}{I_1} \right)$$

or

$$\text{Relative pressure (dB)} = 20 \log \left(\frac{P_2}{P_1} \right)$$

where I_1 and I_2 are intensity values, P_1 and P_2 are pressure amplitude values of sound that are compared as a relative measure, and “log” denotes the base 10 logarithm. In diagnostic ultrasound, the ratio of the intensity of the incident pulse to that of the returning echo can span a range of 1 million times or more! The logarithm function compresses the large and expands the small values into a more manageable number range. An intensity ratio of 10^6 (e.g., an incident intensity 1 million times greater than the returning echo intensity) is equal to 60 dB, whereas an intensity ratio of 10^2 is equal to 20 dB. A change of 10 in the dB scale corresponds to an order of magnitude (ten times) change in intensity; a change of 20 corresponds to two orders of magnitude (100 times) change, and so forth. When the intensity ratio is greater than 1 (e.g., the incident ultrasound intensity to the detected echo intensity), the dB values are positive; when less than 1, the dB values are negative. A loss of 3 dB (−3 dB) represents a 50% loss of signal intensity. The tissue thickness that reduces the ultrasound intensity by 3 dB is considered the “half-value” thickness. Table 16-2 lists a comparison of the dB scale and the corresponding intensity or pressure amplitude ratios.

TABLE 16-2. DECIBEL VALUES, INTENSITY RATIO, AND PRESSURE AMPLITUDE RATIOS

Decibels (dB)	Intensity Ratio		Pressure Amplitude Ratio	
	I_2/I_1	$\text{Log } (I_2/I_1)$	P_2/P_1	$\text{Log } (P_2/P_1)$
0	1	0	1	0
3	2	0.3	1.414	0.15
6	4	0.6	2	0.3
12	16	1.2	4	0.6
20	100	2	10	1
40	10,000	4	100	2
60	1,000,000	6	1000	3
-3	0.5	-0.3	0.707	-0.15
-6	0.25	-0.6	0.5	-0.3
-20	0.01	-2	0.1	-1
-40	0.0001	-4	0.01	-2

Example: Calculate the remaining intensity of a 100-mW ultrasound pulse that loses 30 dB while traveling through tissue.

$$\text{Relative intensity (dB)} = 10 \log \frac{I_2}{I_1}; \quad \text{defining equation}$$

$$-30 \text{ dB} = 10 \log \frac{I_2}{100 \text{ mW}}; \quad \text{divide each side of the equation by 10:}$$

$$-3 = \log \frac{I_2}{100 \text{ mW}}; \quad \text{exponentiate using base 10}$$

$$10^{-3} = \frac{I_2}{100 \text{ mW}}; \quad \text{solve for } I_2:$$

$$I_2 = 0.001 \times 100 \text{ mW} = 0.1 \text{ mW}$$

16.2 INTERACTIONS OF ULTRASOUND WITH MATTER

Ultrasound interactions are determined by the acoustic properties of matter. As ultrasound energy propagates through a medium, interactions that occur include reflection, refraction, scattering, and absorption. Reflection occurs at tissue boundaries where there is a difference in the acoustic impedance of adjacent materials. When the incident beam is perpendicular to the boundary, a portion of the beam (an echo) returns directly back to the source, and the transmitted portion of the beam continues in the initial direction. Refraction describes the change in direction of the transmitted ultrasound energy with nonperpendicular incidence. Scattering occurs by reflection or refraction, usually by small particles within the tissue medium, causes the beam to diffuse in many directions, and gives rise to the characteristic texture and gray scale in the acoustic image. Absorption is the process whereby acoustic energy is converted to heat energy. In this situation, sound energy is lost and cannot be recovered. Attenuation refers to the loss of intensity of the ultrasound beam from absorption and scattering in the medium.

TABLE 16-3. ACOUSTIC IMPEDANCE, $Z = \rho c$, FOR AIR, WATER AND SELECTED TISSUES

Tissue	Z (rayls)
Air	0.0004×10^6
Lung	0.18×10^6
Fat	1.34×10^6
Water	1.48×10^6
Kidney	1.63×10^6
Blood	1.65×10^6
Liver	1.65×10^6
Muscle	1.71×10^6
Skull bone	7.8×10^6

Acoustic Impedance

The acoustic impedance (Z) of a material is defined as

$$Z = \rho c$$

where ρ is the density in kg/m^3 and c is the speed of sound in m/sec . The SI units for acoustic impedance are $\text{kg}/(\text{m}^2\text{sec})$ and are often expressed in rayls, where 1 rayl is equal to $1 \text{ kg}/(\text{m}^2\text{sec})$. Table 16-3 lists the acoustic impedances of materials and tissues commonly encountered in medical ultrasonography. In a simplistic way, the acoustic impedance can be likened to the stiffness and flexibility of a compressible medium such as a spring, as mentioned at the beginning of this chapter. When springs with different compressibility are connected together, the energy transfer from one spring to another depends mostly on stiffness. A large difference in the stiffness results in a large reflection of energy, an extreme example of which is a spring attached to a wall. Minor differences in stiffness or compressibility allow the continued propagation of energy, with little reflection at the interface. Sound propagating through a patient behaves similarly. Soft tissue adjacent to air-filled lungs represents a large difference in acoustic impedance; thus, ultrasonic energy incident on the lungs from soft tissue is almost entirely reflected. When adjacent tissues have similar acoustic impedances, only minor reflections of the incident energy occur. Acoustic impedance gives rise to differences in transmission and reflection of ultrasound energy, which is the basis for pulse echo imaging.

Reflection

The reflection of ultrasound energy at a boundary between two tissues occurs because of the differences in the acoustic impedances of the two tissues. The reflection coefficient describes the fraction of sound intensity incident on an interface that is reflected. For perpendicular incidence (Fig. 16-5A), the reflection pressure amplitude coefficient, R_p , is defined as the ratio of reflected pressure, P_r , and incident pressure, P_i , as

$$R_p = \frac{P_r}{P_i} = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

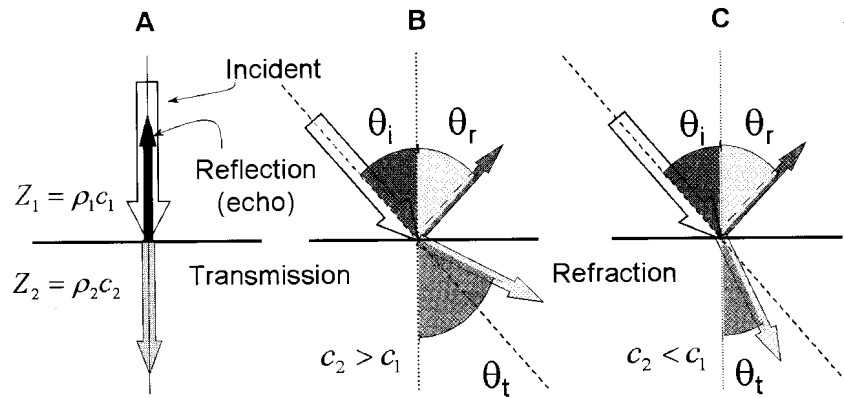


FIGURE 16-5. Reflection and refraction of ultrasound occurs at tissue boundaries with differences in acoustic impedance, Z . With perpendicular incidence (90 degrees) to a tissue boundary, a fraction of the beam is transmitted and a fraction of the beam is reflected back to the source. With nonperpendicular incidence, ($\theta_i \neq 90$ degrees), the reflected fraction of the beam is directed away from the transducer at an angle $\theta_r = \theta_i$, and the transmitted fraction is refracted in the second medium ($\theta_t \neq \theta_i$) when $c_1 \neq c_2$.

The intensity reflection coefficient, R_I , is expressed as the ratio of reflected intensity, I_r , and the incident intensity, I_i , as

$$R_I = \frac{I_r}{I_i} = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2$$

The subscripts 1 and 2 represent tissues proximal and distal to the boundary. The intensity transmission coefficient, T_I , is defined as the fraction of the incident intensity that is transmitted across an interface. With conservation of energy, the intensity transmission coefficient is $T_I = 1 - R_I$. For a fat–muscle interface, the pressure amplitude reflection coefficient and the intensity reflection and transmission coefficients are calculated as

$$R_{P(\text{Fat} \rightarrow \text{Muscle})} = \frac{P_r}{P_i} = \left(\frac{1.71 - 1.34}{1.71 + 1.34} \right) = 0.12;$$

$$R_{I(\text{Fat} \rightarrow \text{Muscle})} = \frac{I_r}{I_i} = \left(\frac{1.71 - 1.34}{1.71 + 1.34} \right)^2 = 0.015; T_{I(\text{Fat} \rightarrow \text{Muscle})} = 1 - R_{I(\text{Fat} \rightarrow \text{Muscle})} = 0.985$$

The actual intensity reflected at a boundary is the product of the incident intensity and the reflection coefficient. For example, an intensity of 40 mW/cm^2 incident on a boundary with $R_I = 0.015$ reflects $40 \times 0.015 = 0.6 \text{ mW/cm}^2$.

The intensity reflection coefficient at a tissue interface is readily calculated from the acoustic impedance of each tissue. Examples of tissue interfaces and respective reflection coefficients are listed in Table 16-4. For a typical muscle–fat interface, approximately 1% of the ultrasound intensity is reflected, and thus almost 99% of the intensity is transmitted to greater depths in the tissues. At a muscle–air interface, nearly 100% of incident intensity is reflected, making anatomy unobservable beyond an air-filled cavity. This is why acoustic coupling gel must be used between the face of the transducer and the skin to eliminate air pockets. A conduit of tissue

TABLE 16-4. PRESSURE AND REFLECTION COEFFICIENTS FOR VARIOUS INTERFACES

Tissue Interface	Pressure Reflection	Intensity Reflection
Liver–kidney	–0.006	0.00003
Liver–fat	–0.10	0.011
Fat–muscle	0.12	0.015
Muscle–bone	0.64	0.41
Muscle–lung	–0.81	0.65
Muscle–air	–0.99	0.99

that allows ultrasound transmission through structures such as the lung is known as an “acoustic window.”

When the beam is perpendicular to the tissue boundary, the sound is returned back to the transducer as an echo. As sound travels from a medium of lower acoustic impedance into a medium of higher acoustic impedance, the reflected wave experiences a 180-degree phase shift in pressure amplitude (note the negative sign on some of the pressure amplitude values in Table 16-4).

The above discussion assumes a “smooth” boundary between tissues, where the wavelength of the ultrasound beam is much greater than the structural variations of the boundary. With higher frequency ultrasound beams, the wavelength becomes smaller, and the boundary no longer appears smooth relative to the wavelength. In this case, returning echoes are diffusely scattered throughout the medium, and only a small fraction of the incident intensity returns to the source (the ultrasound transducer, as described below).

For nonperpendicular incidence at an angle θ_i (Fig. 16-5B,C), the ultrasound energy is reflected at an angle θ_r equal to the incident angle, $\theta_i = \theta_r$. Echoes are directed away from the source of ultrasound, causing loss of the returning signal from the boundary.

Refraction

Refraction describes the change in direction of the transmitted ultrasound energy at a tissue boundary when the beam is not perpendicular to the boundary. Ultrasound frequency does not change when propagating into the next tissue, but a change in the speed of sound may occur. Angles of incidence, reflection, and transmission are measured relative to the normal incidence on the boundary (Fig. 16-5B,C). The angle of refraction (θ_t) is determined by the change in the speed of sound that occurs at the boundary, and is related to the angle of incidence (θ_i) by Snell’s law:

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{c_2}{c_1}$$

where θ_i and θ_t are the incident and transmitted angles, c_1 and c_2 are the speeds of sound in medium 1 and 2, and medium 2 carries the transmitted ultrasound energy. For small angles of incidence and transmission, Snell’s law can be approximated as

$$\frac{\theta_t}{\theta_i} \cong \frac{c_2}{c_1}$$

When $c_2 > c_1$, the angle of transmission is greater than the angle of incidence (Fig. 16-5B), and the opposite with $c_2 < c_1$ (Fig. 16-5C). No refraction occurs when the speed of sound is the same in the two media, or with perpendicular incidence, and thus a straight-line trajectory occurs. This straight-line propagation is assumed in ultrasound machines, and when refraction occurs, it can cause artifacts in the image.

A situation called total reflection occurs when $c_2 > c_1$ and the angle of incidence of the sound beam with a boundary between two media exceeds an angle called the critical angle. In this case, the refracted portion of the beam does not penetrate the second medium at all, but travels along the boundary. The critical angle (θ_c) is calculated by setting $\theta_t = 90$ degrees in Snell's law (equation above) where $\sin(90^\circ) = 1$, producing the equation

$$\sin\theta_c = c_1/c_2$$

Scattering

A specular reflector is a smooth boundary between two media, where the dimensions of the boundary are much larger than the wavelength of the incident ultrasound energy (Fig. 16-6). Acoustic scattering arises from objects within a tissue that are about the size of the wavelength or smaller, and represent a rough or nonspecular reflector surface. Most organs have a characteristic structure that gives rise to a defined scatter "signature" and provides much of the diagnostic information contained in the ultrasound image. Because nonspecular reflectors reflect sound in all directions, the amplitudes of the returning echoes are significantly weaker than echoes from tissue boundaries. Fortunately, the dynamic range of the ultrasound receiver is sufficient to detect echo information over a wide range of amplitudes. In addition, the intensities of returning echoes from nonspecular reflectors in the tissue parenchyma are not greatly affected by beam direction, unlike the strong directional dependence of specular reflectors (Fig. 16-6). Thus, parenchyma-generated

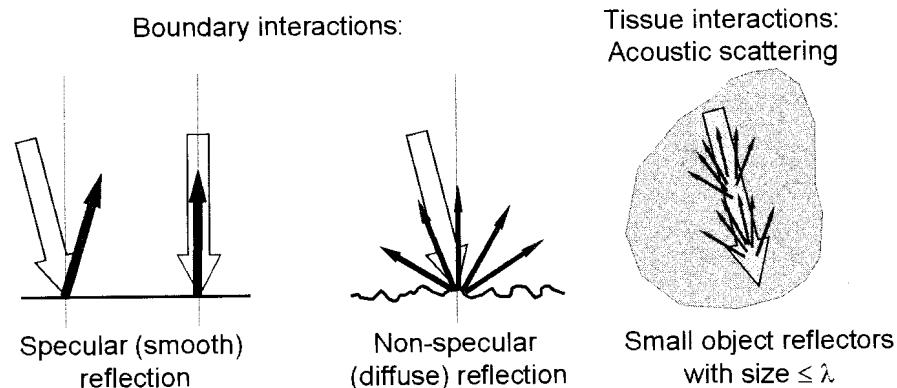


FIGURE 16-6. Ultrasound interactions with boundaries and particles. The characterization of a boundary as specular and nonspecular is partially dependent on the wavelength of the incident ultrasound. As the wavelength becomes smaller, the boundary becomes "rough," resulting in diffuse reflections from the surface because of irregularities. Small particle reflectors within a tissue or organ cause a diffuse scattering pattern that is characteristic of the particle size, giving rise to specific tissue or organ "signatures."

echoes typically have similar echo strengths and gray-scale levels in the image. Differences in scatter amplitude that occur from one region to another cause corresponding brightness changes on the ultrasound display.

In general, the echo signal amplitude from the insonated tissues depends on the number of scatterers per unit volume, the acoustic impedance differences at the scatterer interfaces, the sizes of the scatterers, and the ultrasonic frequency. The terms *hyperechoic* (higher scatter amplitude) and *hypoechoic* (lower scatter amplitude) describe the scatter characteristics relative to the average background signal. Hyperechoic areas usually have greater numbers of scatterers, larger acoustic impedance differences, and larger scatterers. Acoustic scattering from nonspecular reflectors increases with frequency, while specular reflection is relatively independent of frequency; thus, it is often possible to enhance the scattered echo signals over the specular echo signals by using higher ultrasound frequencies.

Attenuation

Ultrasound attenuation, the loss of acoustic energy with distance traveled, is caused chiefly by scattering and tissue absorption of the incident beam. Absorbed acoustic energy is converted to heat in the tissue. The attenuation coefficient, μ , expressed in units of dB/cm, is the relative intensity loss per centimeter of travel for a given medium. Tissues and fluids have widely varying attenuation coefficients, as listed in Table 16-5 for a 1-MHz ultrasound beam. Ultrasound attenuation expressed in dB is approximately proportional to frequency. An approximate rule of thumb for “soft tissue” is 0.5 dB per cm per MHz, or 0.5 (dB/cm)/MHz. The product of the ultrasound frequency (in MHz) with 0.5 (dB/cm)/MHz gives the approximate attenuation coefficient in dB/cm. Thus, a 2-MHz ultrasound beam will have approximately twice the attenuation of a 1-MHz beam; a 10-MHz beam will suffer ten times the attenuation per unit distance. Since the dB scale progresses logarithmically, the beam intensity is exponentially attenuated with distance (Fig. 16-7). The ultrasound half value thickness (HVT) is the thickness of tissue necessary to attenuate the incident intensity by 50%, which is equal to a 3-dB reduction in intensity (6 dB drop in pressure amplitude). As the frequency increases, the HVT decreases, as demonstrated by the examples below.

TABLE 16-5. ATTENUATION COEFFICIENTS FOR SELECTED TISSUES AT 1 MHZ

Tissue Composition	Attenuation coefficient (1 MHz beam, dB/cm)
Water	0.0002
Blood	0.18
Soft tissues	0.3–0.8
Brain	0.3–0.5
Liver	0.4–0.7
Fat	0.5–1.8
Smooth muscle	0.2–0.6
Tendon	0.9–1.1
Bone, cortical	13–26
Lung	40

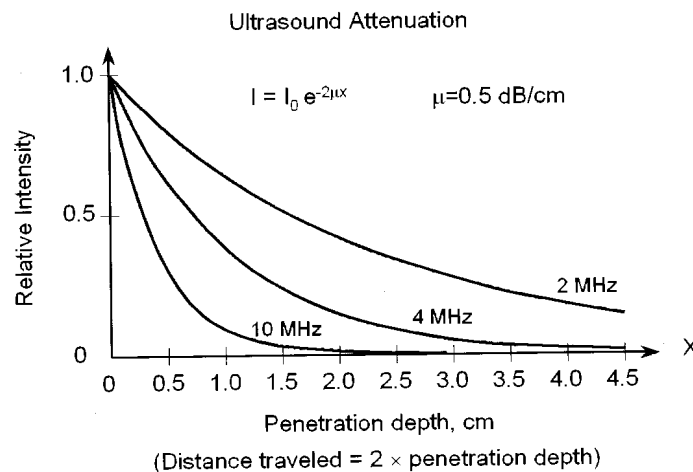


FIGURE 16-7. Ultrasound attenuation occurs exponentially with penetration depth, and increases with increased frequency. Each curve shows intensity of ultrasound of a particular frequency as a function of penetration depth in a medium whose attenuation coefficient is (0.5 dB/cm)/MHz. Note that the total distance traveled by the ultrasound pulse and echo is twice the penetration depth.

Example: Calculate the approximate intensity HVT in soft tissue for ultrasound beams of 2 MHz and 10 MHz. Determine the number of HVTs the incident beam and the echo travel at a 6-cm depth.

Answer: Information needed is (a) the attenuation coefficient approximation 0.5 (dB/cm)/MHz, and (b) one HVT produces a 3-dB loss. Given this information, the HVT in soft tissue for a f MHz beam is

$$\text{HVT}_{f\text{MHz}} (\text{cm}) = \frac{3 \text{ dB}}{\text{Attenuation coefficient (dB/cm)}} = \frac{3 \text{ dB}}{\frac{0.5(\text{dB/cm})}{\text{MHz}} \times f \text{ MHz}} = \frac{6}{f}$$

$$\text{HVT}_{2\text{MHz}} = \frac{3 \text{ dB}}{\frac{0.5(\text{dB/cm})}{\text{MHz}} \times 2 \text{ MHz}} = 3 \text{ cm}$$

$$\text{HVT}_{10\text{MHz}} = \frac{6}{10} = 0.6 \text{ cm}$$

Number of HVTs:

A 6-cm depth requires a travel distance of 12 cm (round trip).

For a 2-MHz beam, this is $12 \text{ cm} / (3 \text{ cm} / \text{HVT}_{2\text{MHz}}) = 4 \text{ HVT}_{2\text{MHz}}$.

For a 10-MHz beam this is $12 \text{ cm} / (0.6 \text{ cm} / \text{HVT}_{10\text{MHz}}) = 20 \text{ HVT}_{10\text{MHz}}$.

Example: Calculate the approximate intensity loss of a 5-MHz ultrasound wave traveling round trip to a depth of 4 cm in liver and reflected from an encapsulated air pocket (100% reflection at the boundary).

Answer: Using 0.5 dB/(cm-MHz) for a 5-MHz transducer, the attenuation coefficient is 2.5 dB/cm. The total distance traveled by the ultrasound pulse is 8 cm (4 cm to the depth of interest and 4 cm back to the transducer). Thus, the total attenuation is 2.5 dB/cm $\times$ 8 cm = 20 dB. The incident intensity relative to the returning intensity (100% reflection at the boundary) is

$$20 \text{ dB} = 10 \log \left(\frac{I_{\text{Incident}}}{I_{\text{Echo}}} \right)$$

$$2 = \log \left(\frac{I_{\text{Incident}}}{I_{\text{Echo}}} \right)$$

$$10^2 = \frac{I_{\text{Incident}}}{I_{\text{Echo}}}; \text{ therefore, } I_{\text{Incident}} = 100 I_{\text{Echo}}; I_{\text{Echo}} = 0.01 I_{\text{Incident}}$$

The echo intensity is one hundredth of the incident intensity in this example (20 dB). If the boundary reflected just 1% of the incident intensity (a typical value), the returning echo intensity would be or 10,000 times less than the incident intensity (40 dB). Considering the depth and travel distance of the ultrasound energy, the detector system must have a dynamic range of 120 to 140 dB in pressure amplitude variations (up to a 10,000,000 times range) to be sensitive to acoustic signals generated in the medium. When penetration to deeper structures is important, lower frequency ultrasound transducers must be used, because of the strong dependence of attenuation with frequency. Another consequence of frequency-dependent attenuation is the preferential removal of the highest frequency components in a broadband ultrasound pulse (discussed below in section 16.3), and a shift to lower frequencies.

16.3 TRANSDUCERS

Ultrasound is produced and detected with a transducer, composed of one or more ceramic elements with electromechanical properties. The ceramic element converts electrical energy into mechanical energy to produce ultrasound and mechanical energy into electrical energy for ultrasound detection. Over the past several decades, the transducer assembly has evolved considerably in design, function, and capability, from a single-element resonance crystal to a broadband transducer array of hundreds of individual elements. A simple single-element, plane-piston source transducer is illustrated in Fig. 16-8. Major components include the piezoelectric material, matching layer, backing block, acoustic absorber, insulating cover, sensor electrodes, and transducer housing.

Piezoelectric Materials

A piezoelectric material (often a crystal or ceramic) is the functional component of the transducer. It converts electrical energy into mechanical (sound) energy by physical deformation of the crystal structure. Conversely, mechanical pressure applied to its surface creates electrical energy. Piezoelectric materials are characterized by a

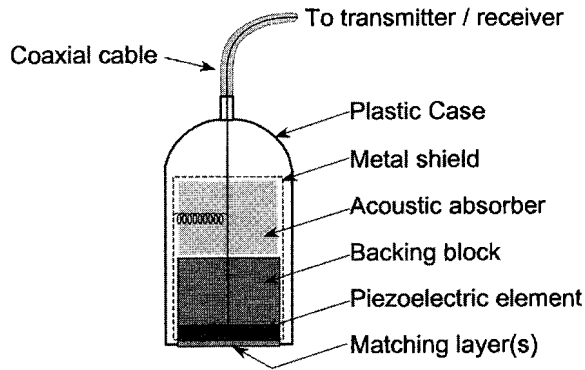


FIGURE 16-8. A single element ultrasound transducer assembly is composed of the piezoelectric ceramic, the backing block, acoustic absorber, metal shield, transducer housing, coaxial cable and voltage source, and the ceramic-to-tissue matching layer.

well-defined molecular arrangement of electrical dipoles (Fig. 16-9). An electrical dipole is a molecular entity containing positive and negative electric charges that has no net charge. When mechanically compressed by an externally applied pressure, the alignment of the dipoles is disturbed from the equilibrium position to cause an imbalance of the charge distribution. A potential difference (voltage) is created across the element with one surface maintaining a net positive charge and one surface a net negative charge. Surface electrodes measure the voltage, which is proportional to the incident mechanical pressure amplitude. Conversely, application of an external voltage through conductors attached to the surface electrodes induces the mechanical expansion and contraction of the transducer element.

There are natural and synthetic piezoelectric materials. An example of a natural piezoelectric material is quartz crystal, commonly used in watches and other timepieces to provide a mechanical vibration source at 32.768 kHz for interval timing. (This is one of several oscillation frequencies of quartz, determined by the crystal cut and machining properties.) Ultrasound transducers for medical imaging applications employ a synthetic piezoelectric ceramic, most often lead-zirconate-titanate (PZT). The piezoelectric attributes are attained after a process of molecular synthesis, heating, orientation of internal dipole structures with an applied external voltage, cooling to permanently maintain the dipole orientation, and cutting into a specific shape. For PZT in its natural state, no piezoelectric properties are exhibited; however, heating the material past its “Curie temperature” (i.e., 328° to 365°C) and applying an external voltage causes the dipoles to align in the ceramic. The external voltage is maintained until the material has cooled to below its Curie temperature. Once the material has cooled, the dipoles retain their alignment. At equilibrium, there is no net charge on ceramic surfaces. When compressed, an imbalance of charge produces a voltage between the surfaces. Similarly, when a voltage is applied between electrodes attached to both surfaces, mechanical deformation occurs (Fig. 16-9B).

Resonance Transducers

Resonance transducers for pulse echo ultrasound imaging are manufactured to operate in a “resonance” mode, whereby a voltage (commonly 150 V) of very short duration (a voltage spike of $\sim 1 \mu\text{sec}$) is applied, causing the piezoelectric material

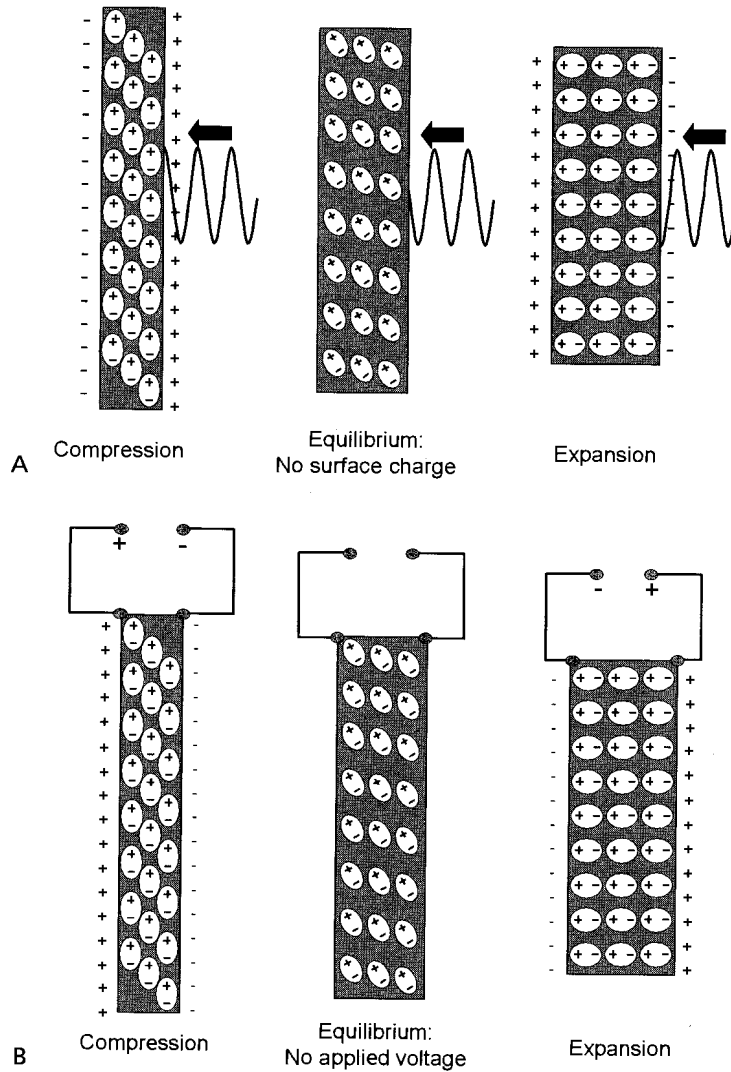


FIGURE 16-9. The piezoelectric element is composed of aligned molecular dipoles. **A:** Under the influence of mechanical pressure from an adjacent medium (e.g., an ultrasound echo), the element thickness contracts (at the peak pressure amplitude), achieves equilibrium (with no pressure) or expands (at the peak rarefactional pressure), causing realignment of the electrical dipoles to produce positive and negative surface charge. Surface electrodes (not shown) measure the voltage as a function of time. **B:** An external voltage source applied to the element surfaces causes compression or expansion from equilibrium by realignment of the dipoles in response to the electrical attraction or repulsion force.

f_0 is determined by the transducer thickness equal to $\frac{1}{2} \lambda$.

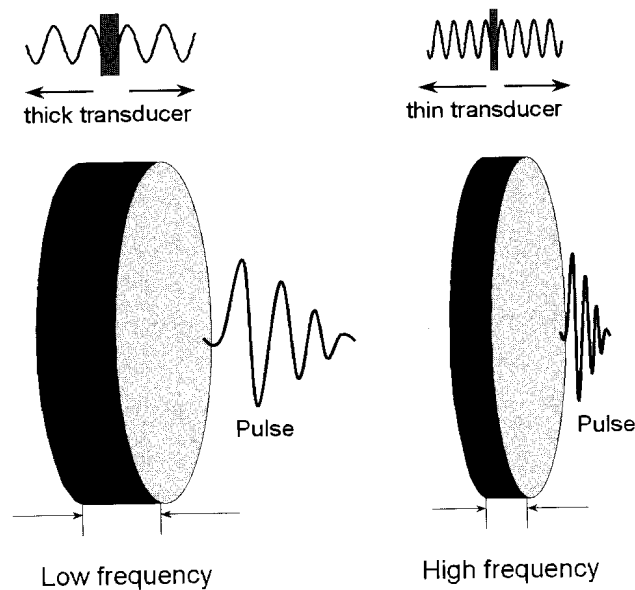


FIGURE 16-10. A short duration voltage spike causes the resonance piezoelectric element to vibrate at its natural frequency, f_0 , which is determined by the thickness of the transducer equal to $\frac{1}{2}\lambda$. Lower-frequency oscillation is produced with a thicker piezoelectric element. The spatial pulse length is a function of the operating frequency and the adjacent damping block.

to initially contract, and subsequently vibrate at a natural resonance frequency. This frequency is selected by the “thickness cut,” due to the preferential emission of ultrasound waves whose wavelength is twice the thickness of the piezoelectric material (Fig. 16-10). The operating frequency is determined from the speed of sound in, and the thickness of, the piezoelectric material. For example, a 5-MHz transducer will have a wavelength in PZT (speed of sound in PZT is $\sim 4,000$ m/sec) of

$$\lambda = \frac{c}{f} = \frac{4,000 \text{ m/sec}}{5 \times 10^6/\text{sec}} = 8 \times 10^{-4} \text{ m} = 0.80 \text{ mm}$$

To achieve the 5-MHz resonance frequency, a transducer element thickness of $\frac{1}{2} \times 0.8 \text{ mm} = 0.4 \text{ mm}$ is required. Higher frequencies are achieved with thinner elements, and lower frequencies with thicker elements. Resonance transducers transmit and receive preferentially at a single “center frequency.”

Damping Block

The damping block, layered on the back of the piezoelectric element, absorbs the backward directed ultrasound energy and attenuates stray ultrasound signals from the housing. This component also dampens the transducer vibration to create an ultrasound pulse with a short spatial pulse length, which is necessary to preserve detail along the beam axis (axial resolution). Dampening of the vibration (also known as “ring-down”) lessens the purity of the resonance frequency and introduces a broadband frequency spectrum. With ring-down, an increase in the bandwidth (range of frequencies) of the ultrasound pulse occurs by introducing higher and lower frequencies above and below the center (resonance) frequency. The “Q factor” describes the bandwidth of the sound emanating from a transducer as

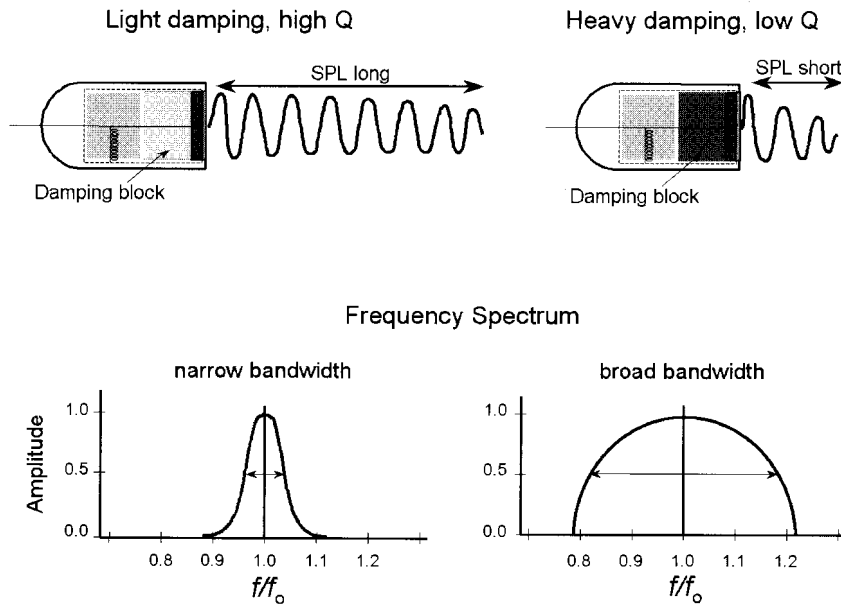


FIGURE 16-11. Effect of damping block on the frequency spectrum. The damping block is adjacent to the back of the transducer and limits the vibration of the element to a small number of cycles. Light damping allows many cycles to occur, which results in an extended spatial pulse length (number of cycles times the wavelength) and a narrow frequency bandwidth (range of frequencies contained in the pulse). Heavy damping reduces the spatial pulse length and broadens the frequency bandwidth. The Q factor is the center frequency divided by the bandwidth.

$$Q = \frac{f_0}{\text{Bandwidth}}$$

where f_0 is the center frequency and the bandwidth is the width of the frequency distribution.

A “high Q ” transducer has a narrow bandwidth (i.e., very little damping) and a corresponding long spatial pulse length. A “low Q ” transducer has a wide bandwidth and short spatial pulse length. Imaging applications require a broad bandwidth transducer in order to achieve high spatial resolution along the direction of beam travel. Blood velocity measurements by Doppler instrumentation (explained in section 16.9) require a relatively narrow-band transducer response in order to preserve velocity information encoded by changes in the echo frequency relative to the incident frequency. Continuous-wave ultrasound transducers have a very high Q characteristic. An example of a “high Q ” and “low Q ” ultrasound pulse illustrates the relationship to spatial pulse length (Fig. 16-11). While the Q factor is derived from the term *quality factor*, a transducer with a low Q does not imply poor quality in the signal.

Matching Layer

The matching layer provides the interface between the transducer element and the tissue and minimizes the acoustic impedance differences between the transducer and the patient. It consists of layers of materials with acoustic impedances that are

intermediate to those of soft tissue and the transducer material. The thickness of each layer is equal to one-fourth the wavelength, determined from the center operating frequency of the transducer and speed of sound in the matching layer. For example, the wavelength of sound in a matching layer with a speed of sound of 2,000 m/sec for a 5-MHz ultrasound beam is 0.4 mm. The optimal matching layer thickness is equal to $\frac{1}{4}\lambda = \frac{1}{4} \times 0.4 \text{ mm} = 0.1 \text{ mm}$. In addition to the matching layers, acoustic coupling gel (with acoustic impedance similar to soft tissue) is used between the transducer and the skin of the patient to eliminate air pockets that could attenuate and reflect the ultrasound beam.

Nonresonance (Broad-Bandwidth) “Multifrequency” Transducers

Modern transducer design coupled with digital signal processing enables “multifrequency” or “multihertz” transducer operation, whereby the center frequency can be adjusted in the transmit mode. Unlike the resonance transducer design, the piezoelectric element is intricately machined into a large number of small “rods,” and

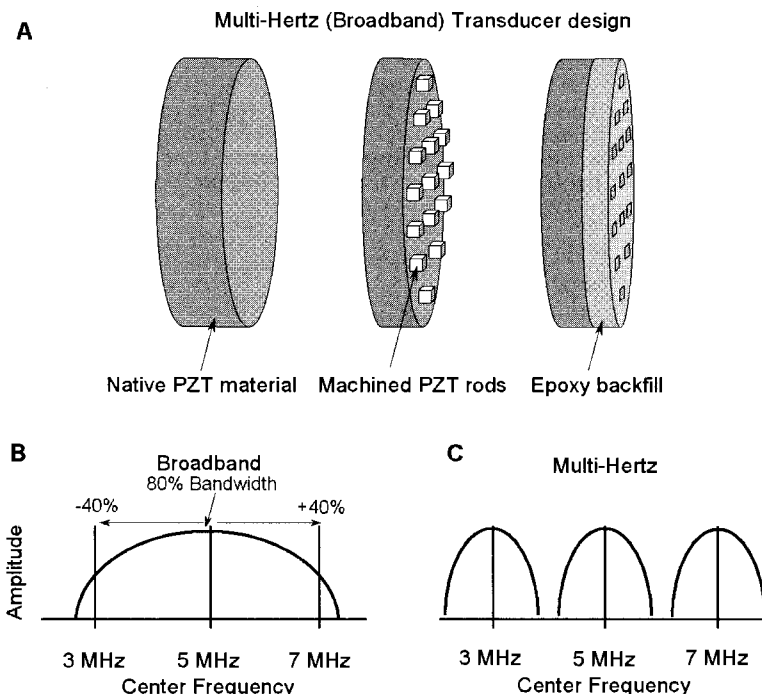


FIGURE 16-12. A: “Multihertz” broadband transducer elements are crafted from a native piezoelectric material into multiple small “rods” with epoxy backfill. This creates a ceramic element with acoustic characteristics closer to soft tissue and produces frequencies with very broad bandwidth. **B:** Bandwidth is often described as a percentage of the center frequency. The graph shows an 80% bandwidth ($\pm 40\%$) for a 5-MHz frequency transducer (3- to 7-MHz operational sensitivity). **C:** Multihertz operation requires the broad bandwidth transducer element to be sensitive to a range of returning frequencies during the reception of echoes with subsequent digital signal processing to select the bandwidth(s) of interest.

then filled with an epoxy resin to create a smooth surface (Fig. 16-12). The acoustic properties are closer to tissue than a pure PZT material, and thus provide a greater transmission efficiency of the ultrasound beam without resorting to multiple matching layers. Multifrequency transducers have bandwidths that exceed 80% of the center frequency.

Excitation of the multifrequency transducer is accomplished with a short square wave burst of ~ 150 V with one to three cycles, unlike the voltage spike used for resonance transducers. This allows the center frequency to be selected within the limits of the transducer bandwidth. Likewise, the broad bandwidth response permits the reception of echoes within a wide range of frequencies. For instance, ultrasound pulses can be produced at a low frequency, and the echoes received at higher frequency. “Harmonic imaging” is a recently introduced technique that uses this ability; lower frequency ultrasound is transmitted into the patient, and the higher frequency harmonics (e.g., two times the transmitted center frequency) created from the interaction with contrast agents and tissues, are received as echoes. Native tissue harmonic imaging has certain advantages including greater depth of penetration, noise and clutter removal, and improved lateral spatial resolution. Multihertz transducers and harmonic imaging are discussed in section 16.7.

Transducer Arrays

The majority of ultrasound systems employ transducers with many individual rectangular piezoelectric elements arranged in linear or curvilinear arrays. Typically, 128 to 512 individual rectangular elements compose the transducer assembly. Each element has a width typically less than half the wavelength and a length of several millimeters. Two modes of activation are used to produce a beam. These are the “linear” (sequential) and “phased” activation/receive modes (Fig. 16-13).

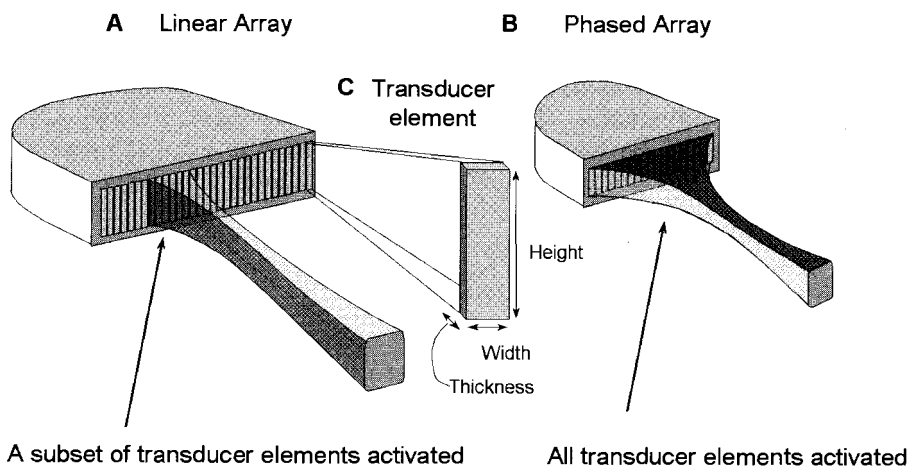


FIGURE 16-13. Multielement transducer arrays. **A:** A linear (or curvilinear) array produces a beam by firing a subset of the total number of transducer elements as a group. **B:** A phased array produces a beam from all of the transducer elements fired with fractional time delays in order to steer and focus the beam. **C:** The transducer element in an array has a thickness, width, and height; the width is typically on the order of $\frac{1}{2}$ wavelength; the height depends on the transducer design.

Linear Arrays

Linear array transducers typically contain 256 to 512 elements; physically these are the largest transducer assemblies. In operation, the simultaneous firing of a small group of ~20 adjacent elements produces the ultrasound beam. The simultaneous activation produces a synthetic aperture (effective transducer width) defined by the number of active elements. Echoes are detected in the receive mode by acquiring signals from most of the transducer elements. Subsequent “A-line” (discussed in section 16.5) acquisition occurs by firing another group of transducer elements displaced by one or two elements. A rectangular field of view is produced with this transducer arrangement. For a curvilinear array, a trapezoidal field of view is produced.

Phased Arrays

A phased-array transducer is usually composed of 64 to 128 individual elements in a smaller package than a linear array transducer. All transducer elements are activated nearly (but not exactly) simultaneously to produce a single ultrasound beam. By using time delays in the electrical activation of the discrete elements across the face of the transducer, the ultrasound beam can be steered and focused electronically without moving the transducer. During ultrasound signal reception, all of the transducer elements detect the returning echoes from the beam path, and sophisticated algorithms synthesize the image from the detected data.

16.4 BEAM PROPERTIES

The ultrasound beam propagates as a longitudinal wave from the transducer surface into the propagation medium, and exhibits two distinct beam patterns: a slightly converging beam out to a distance specified by the geometry and frequency of the transducer (the near field), and a diverging beam beyond that point (the far field). For an unfocused, single-element transducer, the length of the near field is determined by the transducer diameter and the frequency of the transmitted sound (Fig. 16-14). For multiple transducer element arrays, an “effective” transducer diameter is determined by the excitation of a group of transducer elements. Because of the interactions of each of the individual beams and the ability to focus and steer the overall beam, the formulas for a single-element, unfocused transducer are not directly applicable.

The Near Field

The near field, also known as the Fresnel zone, is adjacent to the transducer face and has a converging beam profile. Beam convergence in the near field occurs because of multiple constructive and destructive interference patterns of the ultrasound waves from the transducer surface. Huygen’s principle describes a large transducer surface as an infinite number of point sources of sound energy where each point is characterized as a radial emitter. By analogy, a pebble dropped in a quiet pond creates a radial wave pattern. As individual wave patterns interact, the peaks and troughs from adjacent sources constructively and destructively interfere (Fig. 16-15), causing the beam profile to be tightly collimated in the near field. The ultra-

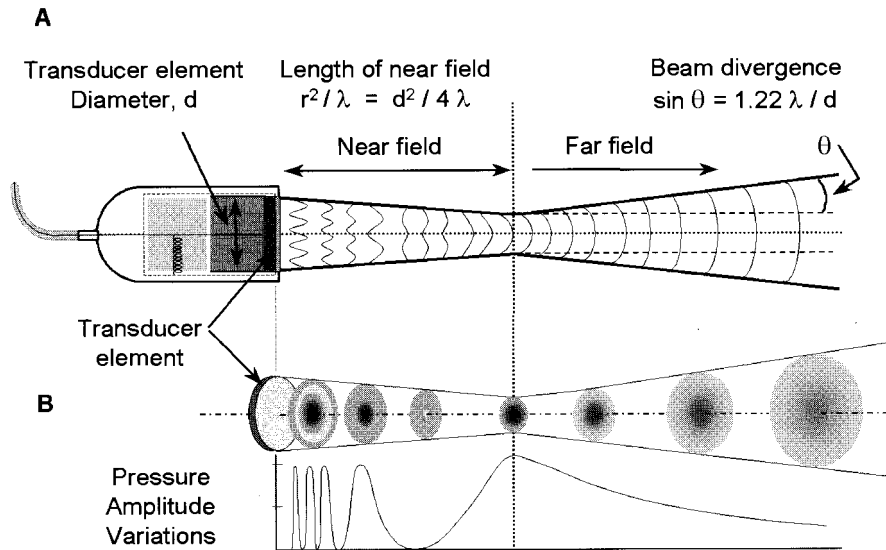


FIGURE 16-14. A: The ultrasound beam from a single, circular transducer element is characterized by the near field and far field regions. **B:** Pressure amplitude variations in the near field are quite complex and rapidly changing, and monotonically decreasing in the far field.

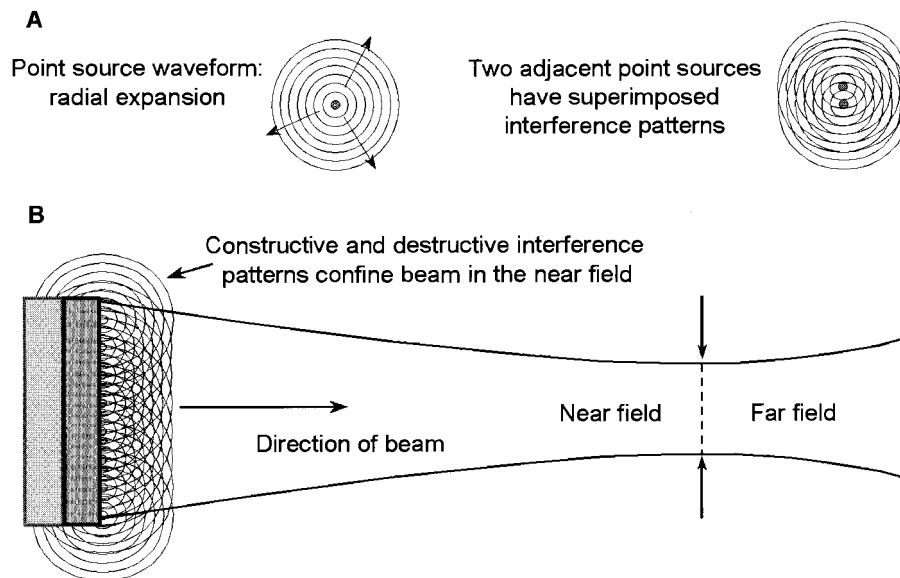


FIGURE 16-15. A: A point source radiates the ultrasound beam as a radially expanding wave. Two adjacent point sources generate constructive and destructive interference patterns. **B:** Huygen's principle considers the plane of the transducer element to be an "infinite" number of point radiators, each emitting a radially expanding wave. The interference pattern results in a converging ultrasound beam in the near field, with subsequent divergence in the far field.

sound beam path is thus largely confined to the dimensions of the active portion of the transducer surface, with the beam diameter converging to approximately half the transducer diameter at the end of the near field. The near field length is dependent on the transducer frequency and diameter:

$$\text{Near field length} = \frac{d^2}{4\lambda} = \frac{r^2}{\lambda}$$

where d is the transducer diameter, r is the transducer radius, and λ is the wavelength of ultrasound in the propagation medium. In soft tissue, $\lambda = \frac{1.54 \text{ mm}}{f(\text{MHz})}$, and the near field length can be expressed as a function of frequency:

$$\text{Near field length, soft tissue (mm)} = \frac{d^2(\text{mm}^2) f(\text{MHz})}{4 \times 1.54 \text{ mm}}$$

A higher transducer frequency (shorter wavelength) will result in a longer near field, as will a larger diameter element (Fig. 16-16). For a 10-mm-diameter transducer, the near field extends 5.7 cm at 3.5 MHz and 16.2 cm at 10 MHz in soft tissue. For a 15-mm-diameter transducer, the corresponding near field lengths are 12.8

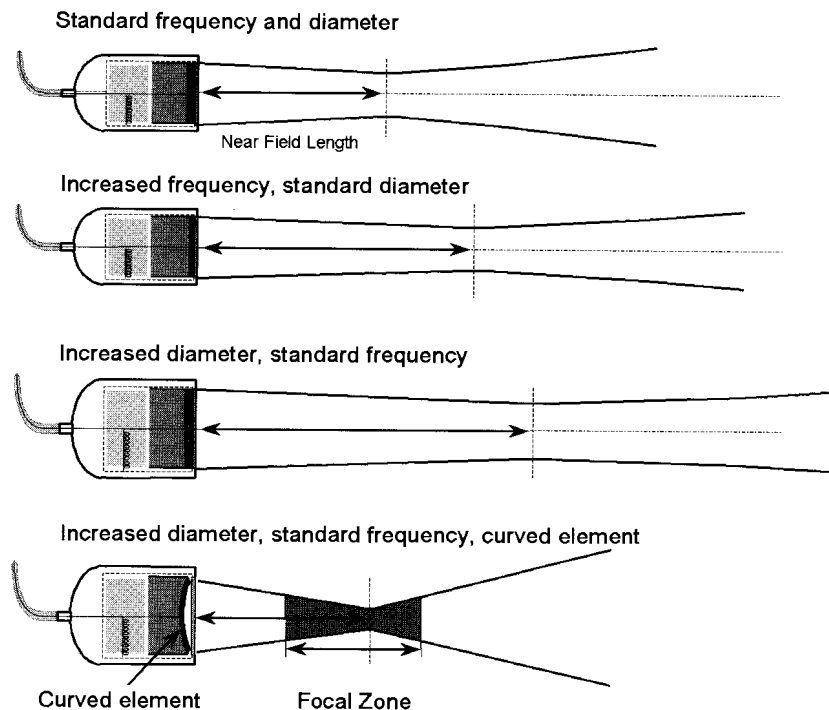


FIGURE 16-16. For an unfocused transducer, the near field length is a function of the transducer frequency and diameter; as either increases, the near field length increases. **Bottom:** A focused single element transducer uses either a curved element (shown) or an acoustic lens. The focal distance (near field length) is brought closer to the transducer surface than the corresponding unfocused transducer, with a decrease in the beam diameter at the focal distance and an increase in the beam divergence angle. A "focal zone" describes the region of best lateral resolution.

and 36.4 cm, respectively. Lateral resolution (the ability of the system to resolve objects in a direction perpendicular to the beam direction) is dependent on the beam diameter and is best at the end of the near field for a single-element transducer. Lateral resolution is worst in areas close to and far from the transducer surface.

Pressure amplitude characteristics in the near field are very complex, caused by the constructive and destructive interference wave patterns of the ultrasound beam. Peak ultrasound pressure occurs at the end of the near field, corresponding to the minimum beam diameter for a single-element transducer. Pressures vary rapidly from peak compression to peak rarefaction several times during transit through the near field. Only when the far field is reached do the ultrasound pressure variations decrease continuously (Fig. 16-14B).

The Far Field

The far field is also known as the Fraunhofer zone, and is where the beam diverges. For a large-area single-element transducer, the angle of ultrasound beam divergence, θ , for the far field is given by

$$\sin \theta = 1.22 \frac{\lambda}{d}$$

where d is the effective diameter of the transducer and λ is the wavelength; both must have the same units of distance. Less beam divergence occurs with high-frequency, large-diameter transducers. Unlike the near field, where beam intensity varies from maximum to minimum to maximum in a converging beam, ultrasound intensity in the far field decreases monotonically with distance.

Focused Transducers

Single-element transducers are focused by using a curved piezoelectric element or a curved acoustic lens to reduce the beam profile (Fig. 16-16, bottom). The focal distance, the length from the transducer to the narrowest beam width, is shorter than the focal length of a nonfocused transducer and is fixed. The focal zone is defined as the region over which the width of the beam is less than two times the width at the focal distance; thus, the transducer frequency and dimensions should be chosen to match the depth requirements of the clinical situation.

Transducer Array Beam Formation and Focusing

In a transducer array, the narrow piezoelectric element width (typically less than one wavelength) produces a diverging beam at a distance very close to the transducer face. Formation and convergence of the ultrasound beam occurs with the operation of several or all of the transducer elements at the same time. Transducer elements in a linear array that are fired simultaneously produce an effective transducer width equal to the sum of the widths of the individual elements. Individual beams interact via constructive and destructive interference to produce a collimated beam that has properties similar to the properties of a single transducer of the same size. With a phased-array transducer, the beam is formed by interaction of the individual wave fronts from each transducer, each with a slight difference in excitation time. Minor

phase differences of adjacent beams form constructive and destructive wave summations that steer or focus the beam profile.

Transmit Focus

For a single transducer or group of simultaneously fired elements in a linear array, the focal distance is a function of the transducer diameter (or the width of the group of simultaneously fired elements), the center operating frequency, and the presence of any acoustic lenses attached to the element surface. Phased array transducers and many linear array transducers allow a selectable focal distance by applying specific timing delays between transducer elements that cause the beam to converge at a specified distance. A shallow focal zone (close to the transducer surface) is produced by firing outer transducers in the array before the inner transducers in a symmetrical pattern (Fig. 16-17). Greater focal distances are achieved by reducing the delay time differences among the transducer elements, resulting in more distal beam convergence. Multiple transmit focal zones are created by repeatedly acquiring data over the same volume, but with different phase timing of the transducer array elements.

Receive Focus

In a phased array transducer, the echoes received by all of the individual transducer elements are summed together to create the ultrasound signal from a given depth. Echoes received at the edge of the element array travel a slightly longer distance than those received at the center of the array, particularly at shallow depths. Signals from individual transducer elements therefore must be rephased to avoid a loss of resolution when the individual signals are synthesized into an image. Dynamic

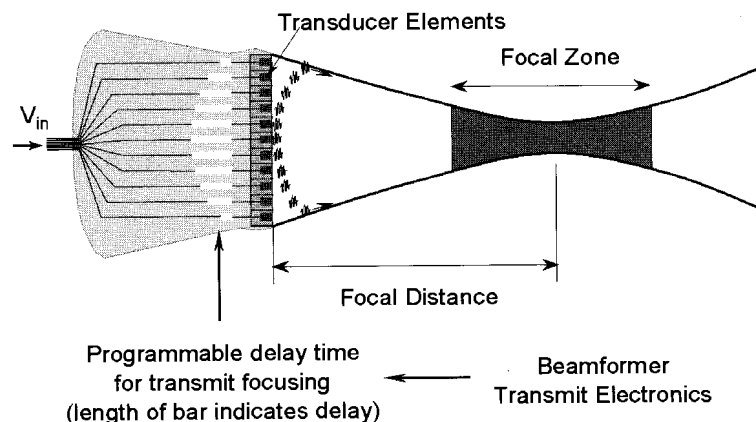


FIGURE 16-17. A phased array transducer assembly uses all elements to produce the ultrasound beam. Focusing is achieved by implementing a programmable delay time (beam-former electronics) for the excitation of the individual transducer elements (focusing requires the outer elements in the array be energized first). Phase differences of the individual ultrasound pulses result in a minimum beam diameter (the focal distance) at a predictable depth.

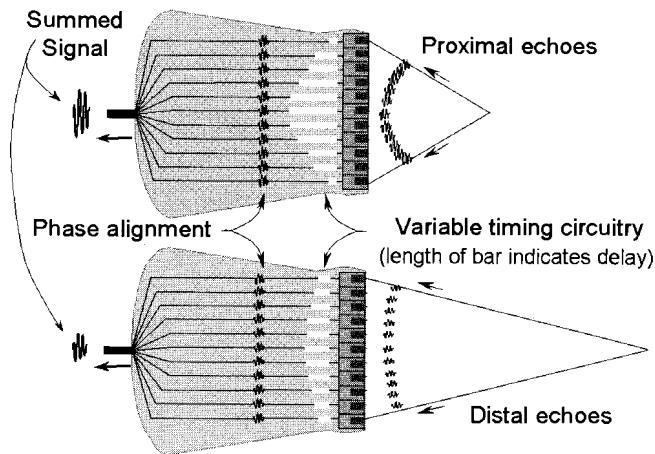


FIGURE 16-18. Dynamic receive focusing. All transducer elements in the phased array are active during the receive mode, and to maintain focus, the receive focus timing must be continuously adjusted to compensate for differences in arrival time across the array as a function of time (depth of the echo). Depicted are an early time (**top**) of proximal echo arrival, and a later time of distal echo arrival. To achieve phase alignment of the echo responses by all elements, variable timing is implemented as a function of element position after the transmit pulse in the beam former. The output of all phase-aligned echoes is summed.

receive focusing is a method to rephase the signals by dynamically introducing electronic delays as a function of depth (time). At shallow depths, rephasing delays between adjacent transducer elements are greatest. With greater depth, there is less phase shift, so the phase delay circuitry for the receiver varies as a function of echo listening time (Fig. 16-18). In addition to phased array transducers, many linear array transducers permit dynamic receive focusing among the active element group.

Dynamic Aperture

The lateral spatial resolution of the linear array beam varies with depth, dependent on the total width of the simultaneously fired elements (aperture). A process termed dynamic aperture increases the number of active receiving elements in the array with reflector depth so that the lateral resolution does not degrade with depth of propagation.

Side Lobes and Grating Lobes

Side lobes are unwanted emissions of ultrasound energy directed away from the main pulse, caused by the radial expansion and contraction of the transducer element during thickness contraction and expansion (Fig. 16-19). In the receive mode of transducer operation, echoes generated from the side lobes are unavoidably remapped along the main beam, which can introduce artifacts in the image. In continuous mode operation, the narrow frequency bandwidth of the transducer (high Q) causes the side lobe energy to be a significant fraction of the total beam. In pulsed mode operation, the low Q broadband ultrasound beam pro-

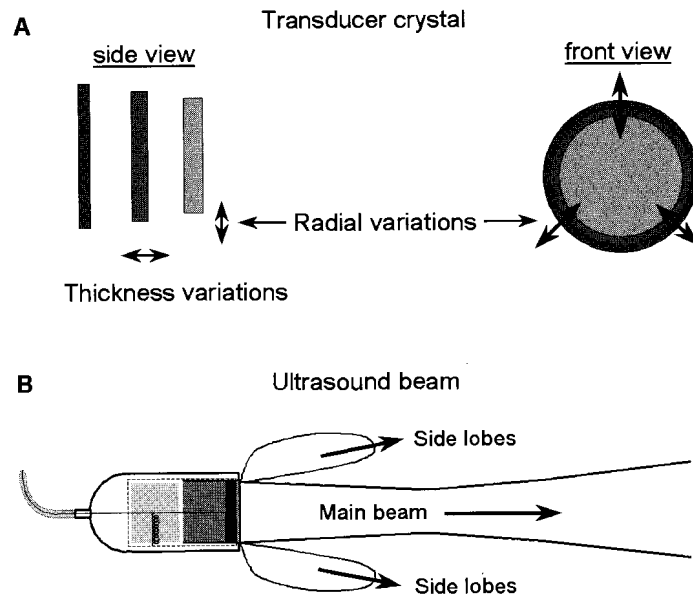


FIGURE 16-19. A: A single-element transducer produces the main ultrasound beam in “thickness mode” vibration; however, radial expansion and contraction also occurs. **B:** Side-lobe energy is created from the radial transducer vibration, which is directed away from the main beam. Echoes received from the side lobe energy are mapped into the main beam, creating unwanted artifacts.

duces a spectrum of acoustic wavelengths that reduces the emission of side lobe energy. For multielement arrays, side lobe emission occurs in a forward direction along the main beam (Fig. 16-20). By keeping the individual transducer element widths small (less than half the wavelength) the side lobe emissions are reduced. Another method to minimize side lobes with array transducers is to reduce the

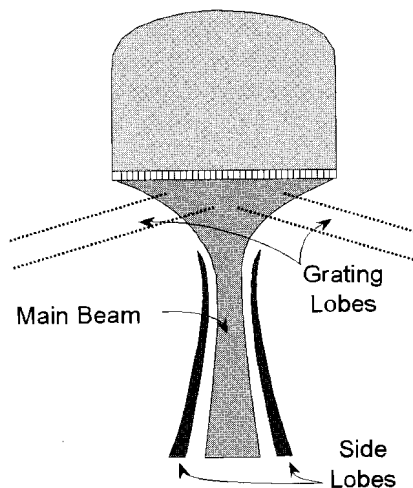


FIGURE 16-20. Side and grating lobes are off-axis energy emissions produced by linear and phased array transducers. Side lobes are forward directed; grating lobes are emitted from the array surface at very large angles.

amplitude of the peripheral transducer element excitations relative to the central element excitations.

Grating lobes result when ultrasound energy is emitted far off-axis by multielement arrays, and are a consequence of the noncontinuous transducer surface of the discrete elements. The grating lobe effect is equivalent to placing a grating in front of a continuous transducer element, producing coherent waves directed at a large angle away from the main beam (Fig. 16-20). This misdirected energy of relatively low amplitude results in the appearance of highly reflective, off-axis objects in the main beam.

Spatial Resolution

In ultrasound, the major factor that limits the spatial resolution and visibility of detail is the volume of the acoustic pulse. The axial, lateral, and elevational (slice thickness) dimensions determine the minimal volume element (Fig. 16-21). Each dimension has an effect on the resolvability of objects in the image.

Axial Resolution

Axial resolution (also known as linear, range, longitudinal, or depth resolution) refers to the ability to discern two closely spaced objects in the direction of the beam. Achieving good axial resolution requires that the returning echoes be distinct without overlap. The minimal required separation distance between two reflectors is one-half of the spatial pulse length (SPL) to avoid the overlap of returning echoes,

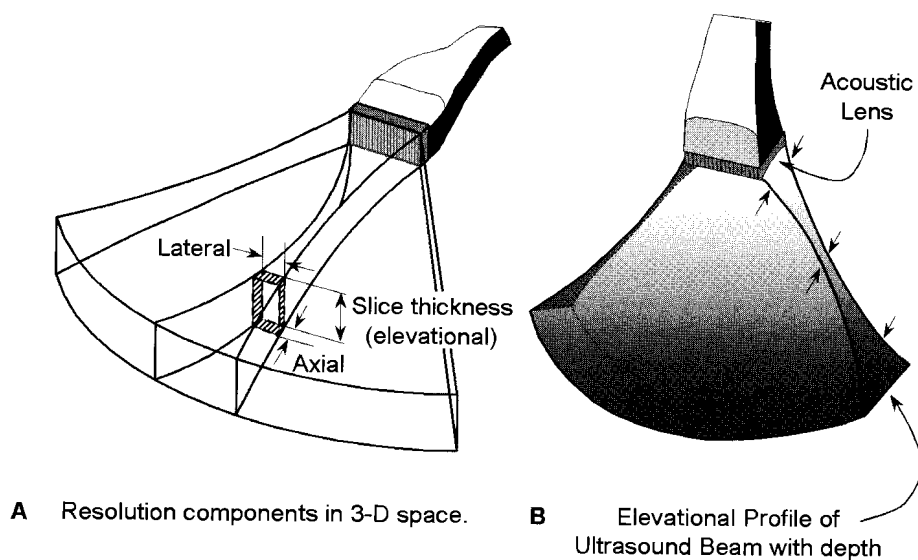


FIGURE 16-21. A: The axial, lateral, and elevational (slice thickness) contributions in three dimensions are shown for a phased array transducer ultrasound beam. Axial resolution, along the direction of the beam, is independent of depth; lateral resolution and elevational resolution are strongly depth dependent. Lateral resolution is determined by transmit and receive focus electronics; elevational resolution is determined by the height of the transducer elements. At the focal distance (see Fig. 16-17), axial is better than lateral and is better than elevational resolution. **B:** Elevational resolution profile with an acoustic lens across the transducer array produces a focal zone in the slice thickness direction.

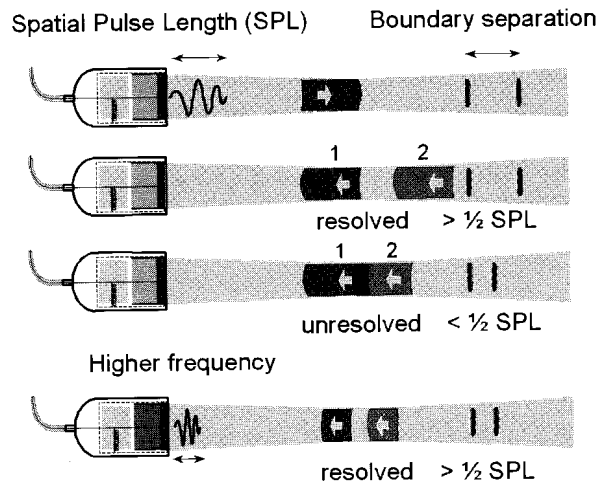


FIGURE 16-22. Axial resolution is equal to $\frac{1}{2}$ spatial pulse length (SPL). Tissue boundaries that are separated by a distance greater than $\frac{1}{2}$ SPL produce echoes from the first boundary that are completely distinct from echoes reflected from the second boundary, whereas boundaries with less than $\frac{1}{2}$ SPL result in overlap of the returning echoes. Higher frequencies reduce the SPL and thus improve the axial resolution (**bottom**).

as the distance traveled between two reflectors is twice the separation distance. Objects spaced closer than $\frac{1}{2}$ SPL will not be resolved (Fig. 16-22).

The SPL is the number of cycles emitted per pulse by the transducer multiplied by the wavelength. Shorter pulses, producing better axial resolution, can be achieved with greater damping of the transducer element (to reduce the pulse duration and number of cycles) or with higher frequency (to reduce wavelength). For imaging applications, the ultrasound pulse typically consists of three cycles. At 5 MHz (wavelength of 0.31 mm), the SPL is about $3 \times 0.31 = 0.93$ mm, which provides an axial resolution of $\frac{1}{2} (0.93 \text{ mm}) = 0.47$ mm. At a given frequency, shorter pulse lengths require heavy damping and low Q, broad-bandwidth operation. For a constant damping factor, higher frequencies (shorter wavelengths) give better axial resolution, but the imaging depth is reduced. Axial resolution remains constant with depth.

Lateral Resolution

Lateral resolution, also known as azimuthal resolution, refers to the ability to discern as separate two closely spaced objects perpendicular to the beam direction. For both single element transducers and multielement array transducers, the beam diameter determines the lateral resolution (Fig. 16-23). Since the beam diameter varies with the distance from the transducer in the near and far field, the lateral resolution is depth dependent. The best lateral resolution occurs at the near field–far field interface. At this depth, the effective beam diameter is approximately equal to half the transducer diameter. In the far field, the beam diverges and substantially reduces the lateral resolution. The typical lateral resolution for an unfocused transducer is approximately 2 to 5 mm. A focused transducer uses an acoustic lens (a curved acoustic material analogous to an optical lens) to decrease the beam diameter at a specified distance from the transducer. With an acoustic lens, lateral resolution at the near field–far field interface is traded for better lateral resolution at a shorter depth, but the far field beam divergence is substantially increased (Fig. 16-16, bottom).

The lateral resolution of linear and curvilinear array transducers can be varied. The number of elements simultaneously activated in a group defines an “effective” transducer width that has similar behavior to a single transducer element of the

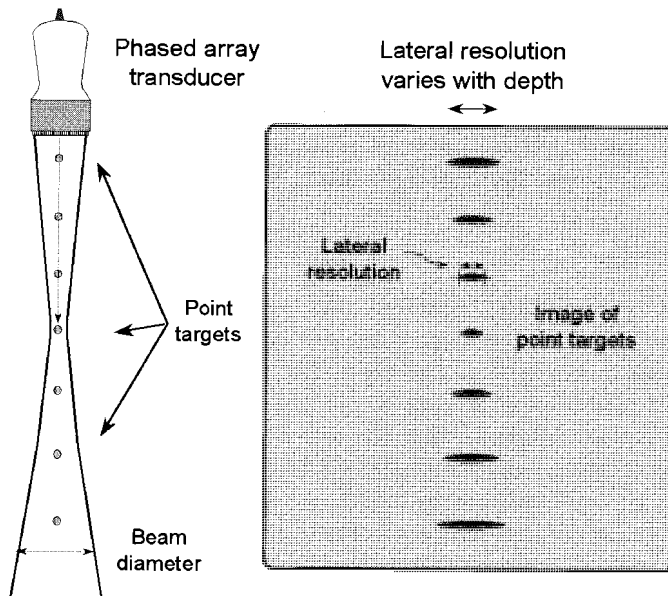


FIGURE 16-23. Lateral resolution indicates the ability to discern objects perpendicular to the direction of beam travel, and is determined by the beam diameter. Point objects in the beam are averaged over the effective beam diameter in the ultrasound image as a function of depth. Best lateral resolution occurs at the focal distance; good resolution occurs over the focal zone.

same width. Transmit and receive focusing can produce focal zones at varying depths along each line.

For the phased array transducer, focusing to a specific depth is achieved by both beam steering and transmit/receive focusing to reduce the effective beam width and improve lateral resolution, especially in the near field. Multiple transmit/receive focal zones can be implemented to maintain lateral resolution as a function of depth (Fig. 16-24). Each focal zone requires separate pulse echo sequences to acquire data.

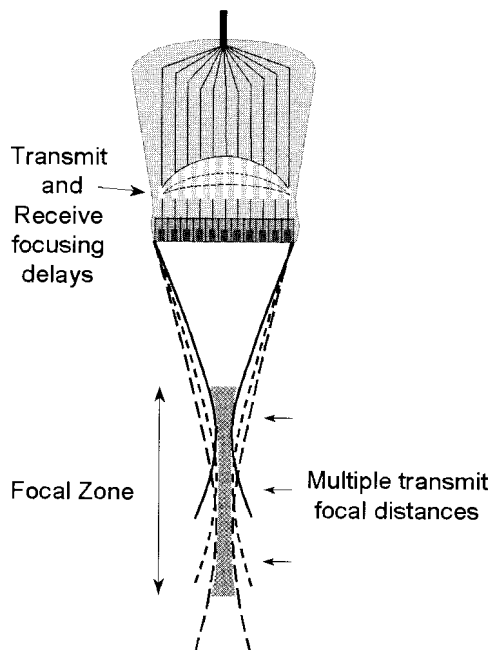


FIGURE 16-24. Phased array transducers permit multiple user selectable transmit and receive focal zones implemented by the beam former electronics. Each focal zone requires the excitation of the entire array for a given focal distance. Good lateral resolution over an extended depth is achieved, but the image rate is reduced.

One way to accomplish this is to acquire data along one beam line multiple times (depending on the number of transmit focal zones), and accept only the echoes within each focal zone, building up a single line of in-focus zones. Increasing the number of focal zones improves overall lateral resolution, but the amount of time required to produce an image increases and reduces the frame rate and/or number of scan lines per image (see Image Quality and Artifacts, below).

Elevational Resolution

The elevational or slice-thickness dimension of the ultrasound beam is perpendicular to the image plane. Slice thickness plays a significant part in image resolution, particularly with respect to volume averaging of acoustic details in the regions close to the transducer and in the far field beyond the focal zone. Elevational resolution is dependent on the transducer element height in much the same way that the lateral resolution is dependent on the transducer element width (Fig. 16-21). Slice thickness is typically the worst measure of resolution for array transducers. Use of a fixed focal length lens across the entire surface of the array provides improved elevational resolution at the focal distance. Unfortunately, this compromises resolution due to partial volume averaging before and after the elevational focal zone (elevational resolution quality control phantom image shows the effects of variable resolution with depth in Fig. 16-55B).

Multiple linear array transducers with five to seven rows, known as 1.5-dimensional (1.5-D) transducer arrays, have the ability to steer and focus the beam in the elevational dimension. Elevational focusing is implemented with phased excitation of the outer to inner arrays to minimize the slice thickness dimension at a given depth (Fig. 16-25). By using subsequent excitations with different focusing distances, multiple transmit focusing can produce smaller slice thickness over a range of tissue depths. A disadvantage of elevational focusing is a frame rate reduction penalty required for multiple excitations to build one image. The increased width of the transducer array also limits positioning flexibility. Extension to full 2D transducer arrays with enhancements in computational power will allow 3D imaging with uniform resolution throughout the image volume.

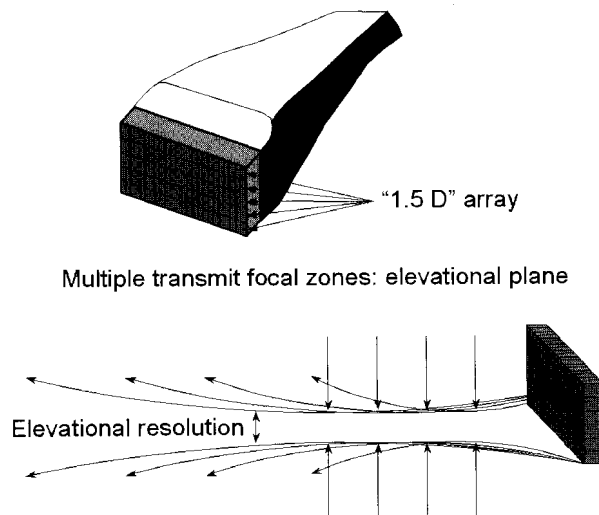


FIGURE 16-25. Elevational resolution with multiple transmit focusing zones is achieved with "1.5-D" transducer arrays to reduce the slice thickness profile over an extended depth. Five to seven discrete arrays replace the single array. Phase delay timing provides focusing in the elevational plane (similar to lateral transmit and receive focusing).

16.5 IMAGE DATA ACQUISITION

Understanding ultrasonic image formation requires knowledge of ultrasound production, propagation, and interactions. Images are created using a pulse echo method of ultrasound production and detection. Each pulse transmits directionally into the patient, and then experiences partial reflections from tissue interfaces that create echoes, which return to the transducer. Image formation using the pulse echo approach requires a number of hardware components: the beam former, pulser, receiver, amplifier, scan converter/image memory, and display system (Fig. 16-26). The detection and processing of the echo signals is the subject of this section.

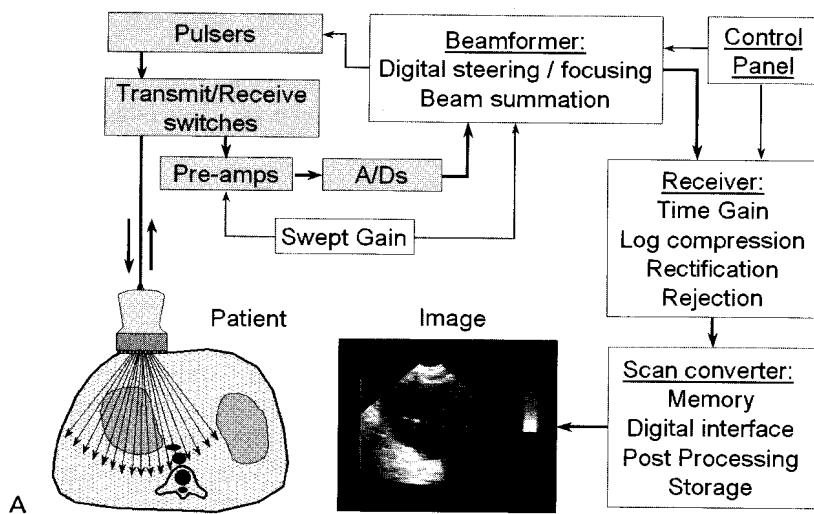


FIGURE 16-26. A: Components of the ultrasound imager. This schematic depicts the design of a digital acquisition/digital beam former system, where each of the transducer elements in the array has a pulser, transmit-receive switch, preamplifier, and analog-to-digital converter (A/D) (e.g., for a 128-element phased array there are 128 components shown as shaded boxes). Swept gain reduces the dynamic range of the signals prior to digitization. The beam former provides focusing, steering, and summation of the beam; the receiver processes the data for optimal display, and the scan converter produces the output image rendered on the monitor. *Thick lines* indicate the path of ultrasound data through the system. **B:** A commercial ultrasound scanner system is composed of a keyboard, various acquisition and processing controls, several transducer selections, an image display monitor, and other components/interfaces (not shown).

Ultrasound equipment is rapidly evolving toward digital electronics and processing, and current state-of-the-art systems use various combinations of analog and digital electronics. The discussion below assumes a hybrid analog and digital processing capability for image data acquisition.

Beam Former

The beam former is responsible for generating the electronic delays for individual transducer elements in an array to achieve transmit and receive focusing and, in phased arrays, beam steering. Most modern, high-end ultrasound equipment incorporates a digital beam former and digital electronics for both transmit and receive functions. A digital beam former controls application-specific integrated circuits (ASICs) that provide transmit/receive switches, digital-to-analog and analog-to-digital converters, and preamplification and time gain compensation circuitry for each of the transducer elements in the array. Each of these components is explained below. Major advantages of digital acquisition and processing include the flexibility to introduce new ultrasound capabilities by programmable software algorithms and to enhance control of the acoustic beam.

Pulser

The pulser (also known as the transmitter) provides the electrical voltage for exciting the piezoelectric transducer elements, and controls the output transmit power by adjustment of the applied voltage. In digital beam-former systems, a digital-to-analog-converter determines the amplitude of the voltage. An increase in transmit amplitude creates higher intensity sound and improves echo detection from weaker reflectors. A direct consequence is higher signal-to-noise ratio in the images, but also higher power deposition to the patient. User controls of the output power are labeled "output," "power," "dB," or "transmit" by the manufacturer. In some systems, a low power setting for obstetric imaging is available to reduce power deposition to the fetus. A method for indicating output power in terms of a thermal index (TI) and mechanical index (MI) is usually provided (see section 16.11).

Transmit/Receive Switch

The transmit/receive switch, synchronized with the pulser, isolates the high voltage used for pulsing (~ 150 V) from the sensitive amplification stages during receive mode, with induced voltages ranging from ~ 1 V to ~ 2 μ V from the returning echoes. After the ring-down time, when vibration of the piezoelectric material has stopped, the transducer electronics are switched to sensing small voltages caused by the returning echoes, over a period up to about 1000 μ sec (1 msec).

Pulse Echo Operation

In the pulse echo mode of transducer operation, the ultrasound beam is intermittently transmitted, with a majority of the time occupied by listening for echoes. The ultrasound pulse is created with a short voltage waveform provided by the pulser of the ultrasound system. This event is sometimes called the main bang. The generated pulse is typically two to three cycles long, dependent on the damping charac-

teristics of the transducer elements. With a speed of sound of 1,540 m/sec (0.154 cm/ μ sec), the time delay between the transmission pulse and the detection of the echo is directly related to the depth of the interface as

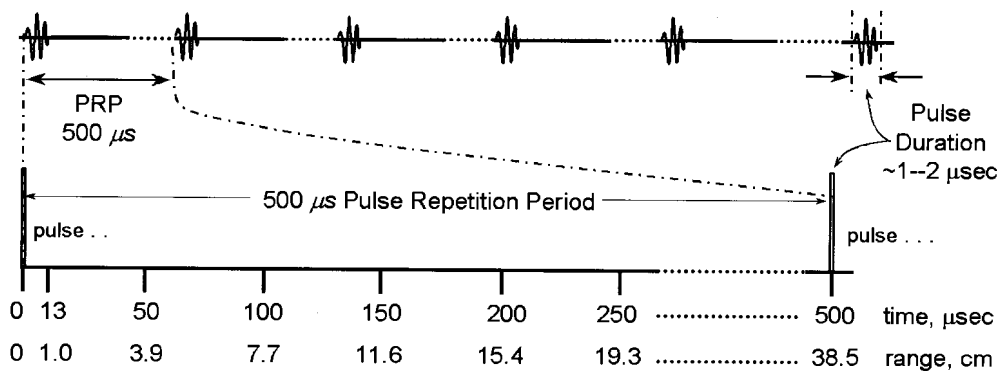
$$\text{Time } (\mu\text{sec}) = \frac{2D(\text{cm})}{c(\text{cm}/\mu\text{sec})} = \frac{2D(\text{cm})}{0.154 \text{ cm}/\mu\text{sec}} = 13 \mu\text{sec} \times D(\text{cm})$$

$$\text{Distance (cm)} = \frac{c(\text{cm}/\mu\text{sec}) \times \text{Time } (\mu\text{sec})}{2} = 0.077 \times \text{Time } (\mu\text{sec})$$

where c , the speed of sound, is expressed in cm/ μ sec; distance from the transducer to the reflector, D , is expressed in cm; the constant 2 represents the round-trip distance; and time is expressed in μ sec. One pulse echo sequence produces one amplitude-modulated (A-line) of image data. The timing of the data excitation and echo acquisition relates to distance (Fig. 16-27). Many repetitions of the pulse echo sequence are necessary to construct an image from the individual A-lines.

The number of times the transducer is pulsed per second is known as the pulse repetition frequency (PRF). For imaging, the PRF typically ranges from 2,000 to 4,000 pulses per second (2 to 4 kHz). The time between pulses is the pulse repetition period (PRP), equal to the inverse of the PRF. An increase in PRF results in a decrease in echo listening time. The maximum PRF is determined by the time required for echoes from the most distant structures to reach the transducer. If a second pulse occurs before the detection of the most distant echoes, these more distant echoes can be confused with prompt echoes from the second pulse, and artifacts can occur. The maximal range is determined from the product of the speed of sound and the PRP divided by 2 (the factor of 2 accounts for round-trip distance):

$$\begin{aligned} \text{Maximal range (cm)} &= 154,000 \text{ cm/sec} \times \text{PRP (sec)} \times 1/2 = 77,000 \times \text{PRP} \\ &= 77,000/\text{PRF} \end{aligned}$$



$$\text{PRF} = \frac{1}{\text{PRP}} = \frac{1}{500 \mu\text{s}} = \frac{1}{500 \times 10^{-6} \text{ s}} = \frac{2000}{\text{s}} = 2 \text{ kHz}$$

FIGURE 16-27. This diagram shows the initial pulse occurring in a very short time span, pulse duration of 1 to 2 μ sec, and the time between pulses equal to the pulse repetition period (PRP) of 500 μ sec. The pulse repetition frequency (PRF) is 2,000/sec, or 2 kHz. Range (one-half the round-trip distance) is calculated assuming a speed of sound = 1,540 m/sec.

TABLE 16-6. TYPICAL PRF, PRP, AND DUTY CYCLE VALUES FOR ULTRASOUND OPERATION MODES

Operation Mode	PRF (Hz)	PRP (μ sec)	Duty Cycle (%)
M-mode	500	2,000	0.05
Real-time	2,000–4,000	500–250	0.2–0.4
Pulsed Doppler	4,000–12,000	250–83	0.4–1.2

Adapted from Zagzebski J. *Essentials of ultrasound physics*. St. Louis: Mosby-Year Book, 1996.

A 500 μ sec PRP corresponds to a PRF of 2 kHz and a maximal range of 38.5 cm. For a PRP of 250 μ sec (PRF of 4 kHz), the maximum depth is halved to 19.3 cm. Higher ultrasound frequency operation has limited penetration depth, allowing high PRFs. Conversely, lower frequency operation requires lower PRFs because echoes can return from greater depths. Ultrasound transducer frequency should not be confused with pulse repetition frequency, and the period of the sound wave ($1/f$) should not be confused with the pulse repetition period ($1/\text{PRF}$). The ultrasound frequency is calibrated in MHz, whereas PRF is in kHz, and the period of the ultrasound is measured in microseconds compared to milliseconds for the PRP.

Pulse duration is the ratio of the number of cycles in the pulse to the transducer frequency, and is equal to the instantaneous “on” time. A pulse consisting of two cycles with a center frequency of 2 MHz has a duration of 1 μ sec. Duty cycle, the fraction of “on” time, is equal to the pulse duration divided by the pulse repetition period. For real-time imaging applications, the duty cycle is typically 0.2% to 0.4%, indicating that greater than 99.5% of the scan time is spent “listening” to echoes as opposed to producing acoustic energy. Intensity levels in medical ultrasonography are very low when averaged over time, as is the intensity when averaged over space due to the collimation of the beam.

For clinical data acquisition, a typical range of PRF, PRP, and duty cycle values are listed in Table 16-6.

Preamplification and Analog to Digital Conversion

In multielement array transducers, all preprocessing steps are performed in parallel. Each transducer element produces a small voltage proportional to the pressure amplitude of the returning echoes. An initial preamplification increases the detected voltages to useful signal levels. This is combined with a fixed swept gain (Figs. 16-26 and 16-28), to compensate for the exponential attenuation occurring with distance traveled. Large variations in echo amplitude (voltage produced in the piezoelectric element) with time are reduced from $\sim 1,000,000:1$ or 120 dB, to about 1,000:1 or 60 dB with these preprocessing steps.

Early ultrasound units used analog electronic circuits for all functions, which were susceptible to drift and instability. Even today, the initial stages of the receiver often use analog electronic circuits. Digital electronics were first introduced in ultrasound for functions such as image formation and display. Since then, there has been a tendency to implement more and more of the signal preprocessing functions in digital circuitry, particularly in high-end ultrasound systems.

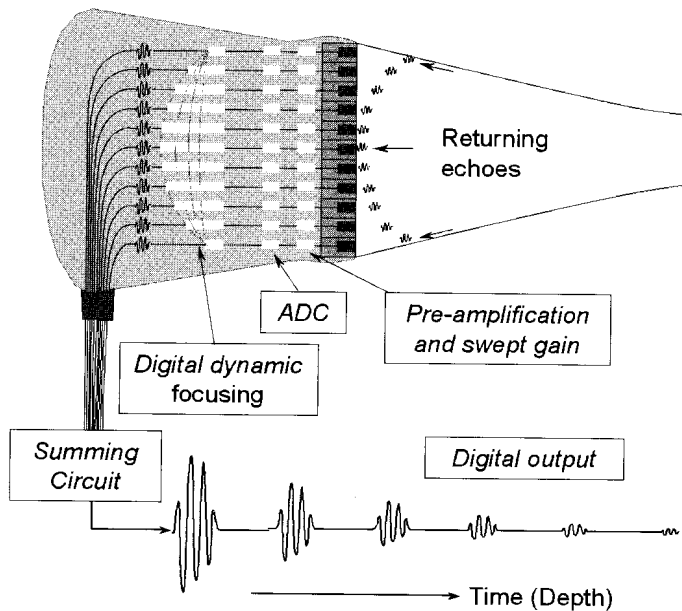


FIGURE 16-28. A phased array transducer produces a pulsed beam that is focused at a programmable depth, and receives echoes during the pulse repetition period. In this digital beam-former system, the analog-to-digital converter (ADC) converts the signals before the beam focus and steering manipulation. Timed electronic delays phase align the echoes, with the output summed to form the ultrasound echo train along a specific beam direction.

In state-of-the-art ultrasound units, each piezoelectric element has its own pre-amplifier and analog-to-digital converter (ADC). A typical sampling rate of 20 to 40 MHz with 8 to 12 bits of precision is used (ADCs are discussed in Chapter 4). ADCs with larger bit depths and sampling rates are necessary for systems that digitize the signals directly from the preamplification stage. In systems where digitization of the signal occurs after analog beam formation and summing, a single ADC with less demanding requirements is typically employed.

Beam Steering, Dynamic Focusing, and Signal Summation

Echo reception includes electronic delays to adjust for beam direction and dynamic receive focusing to align the phases of detected echoes from the individual elements in the array as a function of echo depth. In systems with digital beam formers, this is accomplished with digital processing algorithms. Following phase alignment, the preprocessed signals from all of the active transducer elements are summed. The output signal represents the acoustic information gathered during the pulse repetition period along a single beam direction (Fig. 16-28). This information is sent to the receiver for further processing before rendering into a 2D image.

Receiver

The receiver accepts data from the beam former during the pulse repetition period, which represents echo information as a function of time (depth). Subsequent signal processing occurs in the following sequence (Fig. 16-29):

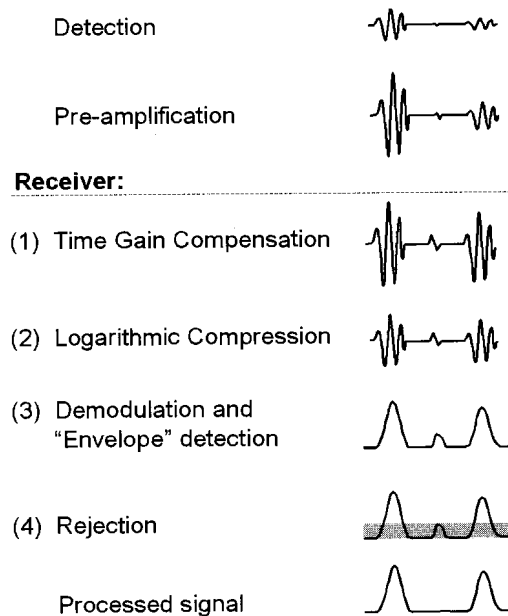


FIGURE 16-29. The receiver processes the data streaming from the beam former. Steps include time gain compensation (TGC), dynamic range compression, rectification, demodulation, and noise rejection. The user has the ability to adjust the TGC and the noise rejection level.

1. Gain adjustments and dynamic frequency tuning. Time gain compensation (TGC) is a user-adjustable amplification of the returning echo signals as a function of time, to further compensate for beam attenuation. TGC (also known as time varied gain, depth gain compensation, and variable swept gain) can be changed to meet the needs of a specific imaging application. The ideal TGC curve makes all equally reflective boundaries equal in signal amplitude, regardless of the depth of the boundary (Fig. 16-30). Variations in the output signals are thus indicative of the acoustic impedance differences between tissue boundaries. User adjustment is typically achieved by multiple slider potentiometers, where each slider represents a given depth in the image, or by a three-knob TGC control, which controls the initial gain, slope, and far gain of the echo signals. For multielement transducers, TGC is applied simultaneously to the signal from each of the individual elements. The TGC amplification further reduces the maximum to minimum range of the echo voltages as a function of time to approximately 50 dB (300:1) from the ~60dB range after preprocessing.

Dynamic frequency tuning is a feature of some broadband receivers that changes the sensitivity of the tuner bandwidth with time, so that echoes from shallow depths are tuned to a higher frequency range, while echoes from deeper structures are tuned to lower frequencies. The purpose of this is to accommodate for beam softening, where increased attenuation of higher frequencies in a broad bandwidth pulse occur as a function of depth. Dynamic frequency tuning allows the receiver to make the most efficient use of the ultrasound frequencies incident on the transducer.

2. Dynamic range compression. Dynamic range defines the effective operational range of an electronic device from the threshold signal level to the saturation level. Key components in the ultrasound detection and display that are most affected by a wide dynamic range include the ADC and the display. For receiver

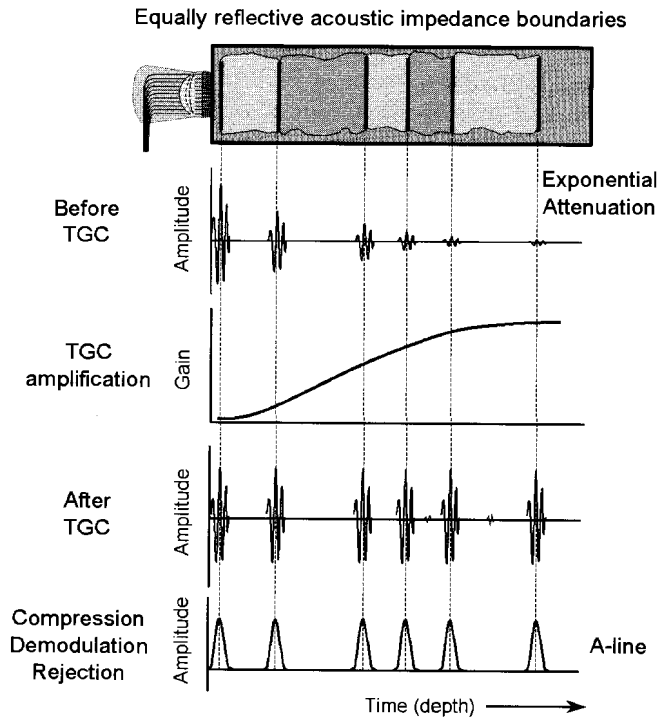


FIGURE 16-30. Time gain compensation amplifies the acquired signals with respect to time after the initial pulse by operator adjustments (usually a set of five to six slide potentiometers; see Fig. 16-26B). Processed A-line data with appropriate TGC demonstrates that equally reflective interfaces have equal amplitude.

systems with an 8-bit ADC, the dynamic range is $20 \log(256) = 48$ dB, certainly not enough to accurately digitize a 50-dB signal. Video monitors and film have a dynamic range of about 100:1 to 150:1, and therefore require a reduced range of input signals to accurately render the display output. Thus, after TGC, the signals must be reduced to 20 to 30 dB, which is accomplished by compression using logarithmic amplification to increase the smallest echo amplitudes and to decrease the largest amplitudes (Fig. 16-31). Logarithmic amplification produces an output signal proportional to the logarithm of the input signal. Logarithmic amplification is performed by an analog signal processor in less costly ultrasound

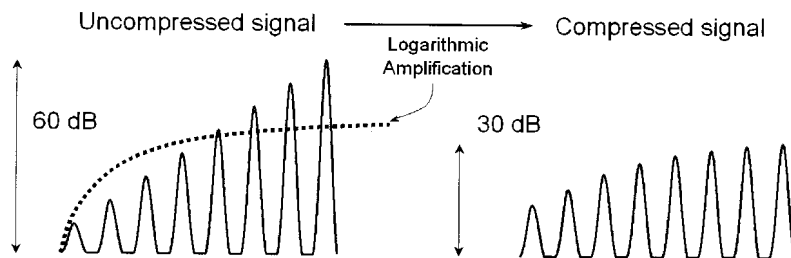


FIGURE 16-31. Logarithmic compression reduces the range of the signals (minimum to maximum) after time gain compensation to enable the full range of gray-scale values to be displayed on the monitor. Shown are the A-line amplitudes reduced by logarithmic amplification from ~60 to ~30 dB.

systems, and digitally in high-end digital systems having ADCs with a large bit depth. In any case, an appropriate range of signals is achieved for display of the amplitude variations as gray scale on the monitor or for printing on film.

3. Rectification, demodulation, and envelope detection. Rectification inverts the negative amplitude signals of the echo to positive values. Demodulation and envelope detection converts the rectified amplitudes of the echo into a smoothed, single pulse.
4. Rejection level adjustment sets the threshold of signal amplitudes allowed to pass to the digitization and display subsystems. This removes a significant amount of undesirable low-level noise and clutter generated from scattered sound or by the electronics.

In the above-listed steps, the operator has the ability to control the time gain compensation, and noise/clutter rejection. The amount of amplification (overall gain) necessary is dependent on the initial power (transmit gain) settings of the ultrasound system. Higher intensities are achieved by exciting the transducer elements with larger voltages. This increases the amplitude of the returning echoes and reduces the need for electronic amplification gain. Conversely, lower ultrasound power settings, such as those used in obstetrical ultrasound, require a greater overall gain to amplify weaker echo signals into the appropriate range for TGC. TGC allows the operator to manipulate depth-dependent gain to improve image uniformity and compensate for unusual imaging situations. Inappropriate adjustment of TGC can lead to artifactual enhancement of tissue boundaries and tissue texture as well as nonuniform response versus depth. The noise rejection level sets a threshold to clean up low-level signals in the electronic signal. It is usually adjusted in conjunction with the transmit power level setting of the ultrasound instrument.

Echo Display Modes

A-Mode

A-mode (A for amplitude) is the display of the processed information from the receiver versus time (after the receiver processing steps shown in Fig. 16-29). As echoes return from tissue boundaries and scatterers (a function of the acoustic impedance differences in the tissues), a digital signal proportional to echo amplitude is produced as a function of time. One "A-line" of data per pulse repetition period is the result. As the speed of sound equates to depth (round-trip time), the tissue interfaces along the path of the ultrasound beam are localized by distance from the transducer. The earliest uses of ultrasound in medicine used A-mode information to determine the midline position of the brain for revealing possible mass effect of brain tumors. A-mode and A-line information is currently used in ophthalmology applications for precise distance measurements of the eye. Otherwise, A-mode display by itself is seldom used.

B-Mode

B-mode (B for brightness) is the electronic conversion of the A-mode and A-line information into brightness-modulated dots on a display screen. In general, the brightness of the dot is proportional to the echo signal amplitude (depending on

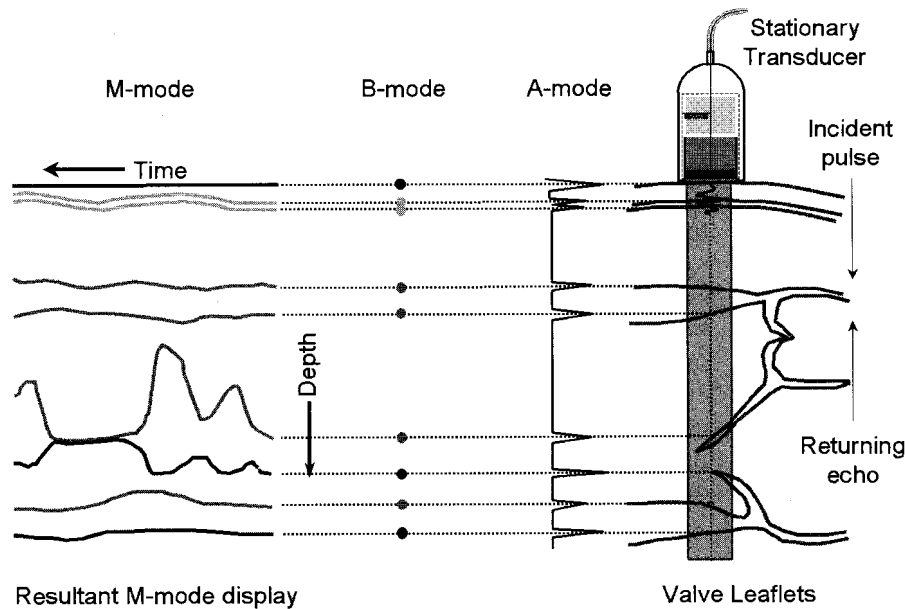


FIGURE 16-32. M-mode data is acquired with a stationary transducer or transducer array. The A-mode data are brightness encoded and deflected across the horizontal direction to produce M-mode curves. M-mode display of valve leaflets is depicted.

signal processing parameters). The B-mode display is used for M-mode and 2D gray-scale imaging.

M-Mode

M-mode (M for motion) is a technique that uses B-mode information to display the echoes from a moving organ, such as the myocardium and valve leaflets, from a fixed transducer position and beam direction on the patient (Fig. 16-32). The echo data from a single ultrasound beam passing through moving anatomy are acquired and displayed as a function of time, represented by reflector depth on the vertical axis (beam path direction) and time on the horizontal axis. M-mode can provide excellent temporal resolution of motion patterns, allowing the evaluation of the function of heart valves and other cardiac anatomy. Only anatomy along a single line through the patient is represented by the M-mode technique, and with advances in real-time 2D echocardiography, Doppler, and color flow imaging, this display mode is of much less importance than in the past.

Scan Converter

The function of the scan converter is to create 2D images from echo information from distinct beam directions, and to perform scan conversion to enable image data to be viewed on video display monitors. Scan conversion is necessary because the image acquisition and display occur in different formats. Early scan converters were of an analog design, using storage cathode ray tubes to capture data. These devices

drifted easily and were unstable over time. Modern scan converters use digital technology for storage and manipulation of data. Digital scan converters are extremely stable, and allow subsequent image processing with application of a variety of mathematical functions.

Digital information streams to the scan converter memory, configured as a matrix of small picture elements (pixels) that represent a rectangular coordinate display. Most ultrasound instruments have a $\sim 500 \times 500$ pixel matrix (variations between manufacturers exist). Each pixel has a memory address that uniquely defines its location within the matrix. During image acquisition, the digital signals are inserted into the matrix at memory addresses that correspond as close as possible to the relative reflector positions in the body. Transducer beam orientation and echo delay times determine the correct pixel addresses (matrix coordinates) in which to deposit the digital information. Misalignment between the digital image matrix and the beam trajectory, particularly for sector-scan formats at larger depths, requires data interpolation to fill in empty or partially filled pixels. The final image is most often recorded with $512 \times 512 \times 8$ bits per pixel, representing about $\frac{1}{4}$ megabyte of data. For color display, the bit depth is often as much as 24 bits (3 bytes per primary color).

16.6 TWO-DIMENSIONAL IMAGE DISPLAY AND STORAGE

A 2D ultrasound image is acquired by sweeping a pulsed ultrasound beam over the volume of interest and displaying echo signals using the B-mode. Echo position is based on the delay time between the pulse initiation and the reception of the echo, using the speed of sound in soft tissue (1,540 m/sec). The 2D image is progressively built up or continuously updated as the beam is swept through the object.

Early B-Mode Scanners—Manual Articulating Arm with Static Display

Early B-mode scanners were made with a single element transducer mounted on an articulating arm with angular position encoders to determine the location of the ultrasound beam path. This information was necessary to place echo signals at proper positions in the image. The mechanical arm constrained the transducer to a plane, so that the resultant image depicted a tomographic slice. An image was built up on an analog scan converter and storage display by repeated pulsing and positional changes of the transducer (Fig. 16-33), requiring several seconds per image. Linear, sector, or “compound” scans could be performed. Compound scans used a combination of linear and sector transducer motions to improve the probability of perpendicular pulse incidence to organ boundaries so that echoes would return to the transducer. Image quality with these systems was highly dependent on the skill of the sonographer.

Mechanical Scanning and Real-Time Display

The next step in the evolution of medical ultrasound imaging was dynamic scanning with “real-time” display. This was achieved by implementing periodic mechanical motion of the transducer. Early scanners used a single element transducer to

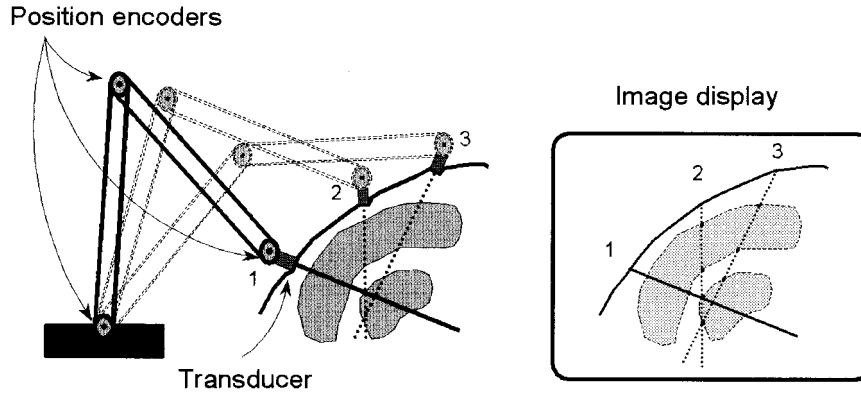


FIGURE 16-33. Articulating arm B-mode scanning produces an acoustic tomographic slice of the body. Position encoders track the three-dimensional position of the transducer to allow the scan converter to place the gray-scale-encoded A-line data in the two-dimensional image plane. Shown is a compound scan (multiple compound directions of the single element transducer) along a plane of the body, and the corresponding image lines.

produce a real-time sector scan image by wobbling the transducer. Enhanced performance was obtained with rotation of a number of transducers on a wheel within the transducer housing (Fig. 16-34). The update of the screen display was determined by the pulse repetition frequency, the oscillation or rotation frequency of the transducer, and the positional information provided by an encoder attached to the rocking or rotating transducer. The number of A-lines that composed the image was determined by the penetration depth and the image update rate.

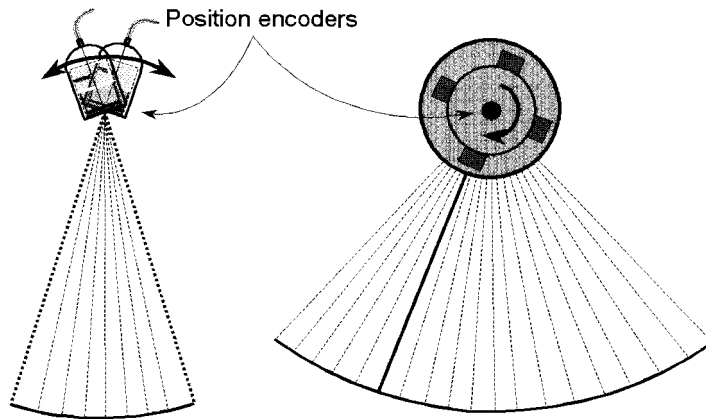


FIGURE 16-34. Dynamic ultrasound imaging with mechanical motion of transducers is accomplished by a back-and-forth wobbling of a single transducer assembly (**left**), or with a rotating multi-element transducer configuration (**right**), among several possible ways. A sector scan image is acquired, composed of a number of A-lines of data. Position encoders provide the trajectory data for the scan converter during the formation of the image.

Electronic Scanning and Real-Time Display

State-of-the-art ultrasound scanners employ array transducers with multiple piezoelectric elements to electronically sweep an ultrasound beam across the volume of interest for dynamic ultrasound imaging. Array transducers are available as linear/curvilinear arrays and phased arrays. They are distinguished by the way in which the beam is produced, and by the field of view coverage that is possible.

Linear and curvilinear array transducers produce rectangular and trapezoidal images, respectively (Fig. 16-35). Such a transducer is typically composed of 256 to 512 discrete transducer elements of less than 1-wavelength width, all in an enclosure from about 6 to 8 cm wide. A small group of adjacent elements (~15 to 20) is simultaneously activated to create an active transducer area defined by the width (sum of the widths of the individual elements in the group) and the height of the elements. This beam propagates perpendicular to the surface of the transducer, with a single line of echo information acquired during the pulse repetition period. A shift of one or more transducer elements and repeating the simultaneous excitation of the group produces the next A-line of data. The ultrasound beam sweeps across the volume of interest in a sequential fashion, with the number of A-lines approximately equal to the number of transducer elements. Advantages of the linear array are the wide field of view (FOV) for regions close to the transducer, and uniform, rectangular sampling across the image. Electronic delays within the subgroup of transducer elements allow transmit and dynamic receive focusing for improved lateral resolution with depth.

The phased array transducer is typically composed of a tightly grouped array of 64, 128, or 256 transducer elements in a 3 to 5 cm wide enclosure. All transducer elements are involved in producing the ultrasound beam and recording the

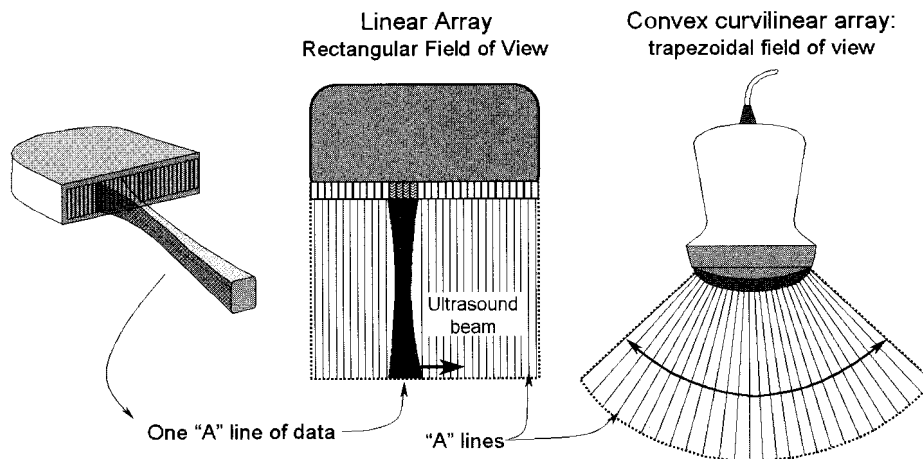


FIGURE 16-35. The linear and curvilinear array transducers produce an image by activating a subgroup of the transducer elements that form one A-line of data in the scanned object, and shifting the active elements by one to acquire the next line of data. Linear arrays produce rectangular image formats; curvilinear arrays produce a trapezoidal format with a wide field of view.

returning echoes. The ultrasound beam is steered by adjusting the delays applied to the individual transducer elements by the beam former. This time delay sequence is varied from one transmit pulse to the next in order to change the sweep angle across the FOV in a sector scan. Figure 16-36 shows three separate beams during the sweep. A similar time-delay strategy (receive focus) is used to spatially synchronize the returning ultrasound echoes as they strike each transducer element (Fig. 16-18). In addition to the beam steering capabilities, lateral focusing, multiple transmit focal zones, and dynamic receive focusing are used in phased arrays.

The small overall size of the phased-array transducer entrance window allows flexibility in positioning, particularly through the intercostal space—a small ultrasound conduit in front of the lungs—to allow cardiac imaging. In addition, it is possible to use the phased array transducer in conjunction with external triggers such as electrocardiogram (ECG) data, to produce M-mode scans at the same time as 2D images, and to allow duplex Doppler/color flow imaging.

Electronic compound scanning is a relatively new static image option that produces a panoramic view of the anatomy. A static image is built up on the display by tracking the position of the array transducer during translation in one direction, and updating the position in the image matrix by the scan converter. This overcomes the limited FOV experienced with sector scans in areas proximal to the transducer, and superimposes multiple views of the anatomy in a single plane.

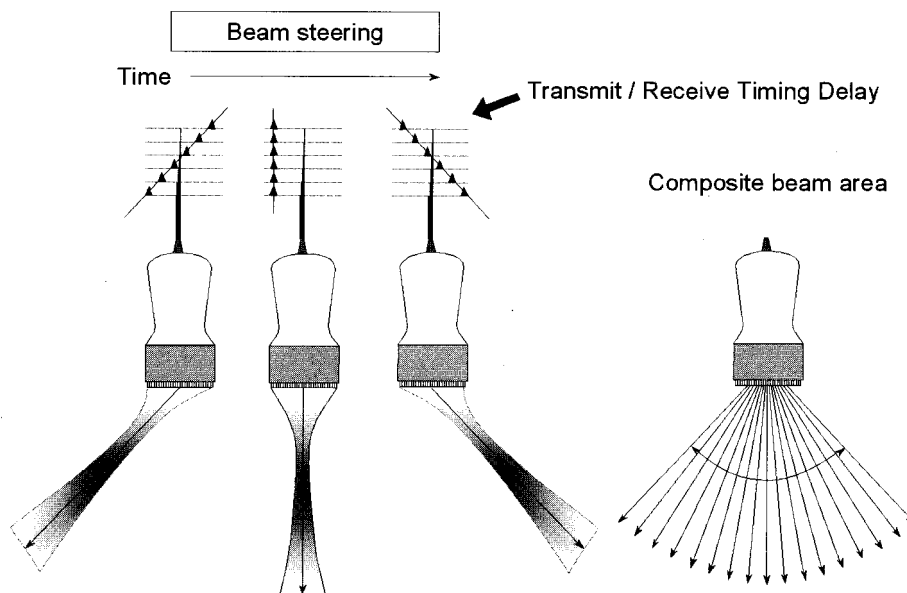


FIGURE 16-36. The beam former electronically steers the ultrasound beam by introducing phase delays during the transmit and receive timing of the phased-array transducer. Lateral focusing also occurs along the beam direction. A sector format composite image is produced (right), with the number of A-lines dependent on several imaging factors discussed in the text.

Image Frame Rate and Spatial Sampling

The 2D image (a single frame) is created from a number of A-lines, N (typically 100 or more), acquired across the field of view. A larger number of lines will produce a higher quality image; however, the finite time for pulse echo propagation places an upper limit on N that depends on the required temporal resolution. The acquisition time for each line, $T_{\text{line}} = 13 \mu\text{sec/cm} \times D$ (cm), is required for the echo data to be unambiguously collected from a depth, D . Thus, the time required per frame, T_{frame} , is given by $N \times T_{\text{line}} = N \times 13 \mu\text{sec/cm} \times D$ (cm). The frame rate per second is the reciprocal of the time required per frame:

$$\text{Frame rate} = \frac{1}{T_{\text{frame}}} = \frac{1}{N \times 13 \mu\text{sec} \times D(\text{cm})} = \frac{0.77/\mu\text{sec}}{N \times D(\text{cm})} = \frac{77,000/\text{sec}}{N \times D(\text{cm})}$$

This equation describes the maximum frame rate possible in terms of N and D . If either N or D increases without a corresponding decrease of the other variable, then the maximum frame rate will decrease. For a given procedure, the ultrasonog-

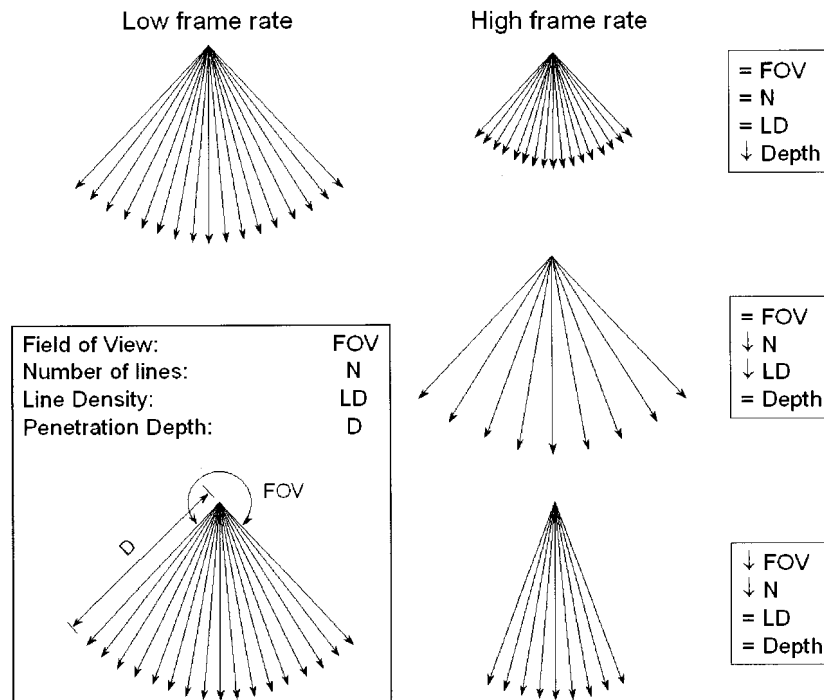


FIGURE 16-37. Ultrasound image quality depends on several factors, including the field of view (FOV), the number of A-lines per image (N), the line density (LD), and penetration depth (D) (inset), in addition to the image frame rate. Compared to the low frame rate acquisition parameters with excellent factors (upper left diagram), a high frame rate (to improve temporal resolution) involves tradeoffs. To maintain a constant FOV, number of A-lines, and line density, penetration depth must be sacrificed (upper right). To maintain penetration depth and FOV, the number of A-lines and line density must be decreased (middle right). To maintain line density and penetration depth, the FOV and number of A-lines must be decreased (bottom right).

rapher must consider the compromises among frame rate, imaging depth, and number of lines/frame. A secondary consideration is the line density (spacing between lines), determined by N and the field of view. Higher frame rates can be achieved by reducing the imaging depth, number of lines, or FOV (Fig. 16-37). The spatial sampling (line density) of the ultrasound beam decreases with depth for sector and trapezoidal scan formats, and remains constant with depth for the rectangular format (linear array).

Another factor that affects frame rate is transmit focusing, whereby the ultrasound beam (each A line) is focused at multiple depths for improved lateral resolution. The frame rate will be decreased by a factor approximately equal to the number of transmit focal zones placed on the image, since the beam-former electronics must transmit an independent set of pulses for each focal zone.

Scan line density is an important component of image quality. Insufficient line density can cause the loss of image resolution, particularly at depth. This might happen when one achieves high temporal resolution (high frame rates) at the expense of line density.

Image Display

For monitor displays, the digital scan converter memory requires a digital-to-analog converter (DAC) and electronics to produce a compatible video signal. The DAC converts the digital pixel data matrix into a corresponding analog video signal compatible with specific video monitors (see Chapter 4 for a detailed explanation). Window and level adjustments of the digital data modify the brightness and contrast of the displayed image without changing the digital memory values (only the look-up transformation table) before digital-to-analog conversion.

The pixel density of the display monitor can limit the quality and resolution of the image. Employing a “zoom” feature on many ultrasound instruments can enlarge the image information to better display details within the image that are otherwise blurred. Two types of methods, “read” zoom, and “write” zoom, are usually available. “Read” zoom enlarges a user-defined region of the stored image and expands the information over a larger number of pixels in the displayed image. Even though the displayed region becomes larger, the resolution of the image itself does not change. Using “write” zoom requires the operator to rescan the area of the patient that corresponds to the user-selected area. When write zoom is enabled, the transducer scans the selected area and only the echo data within the limited region is acquired. This allows a greater line density and all pixels in the memory are used in the display of the information. In this case, better sampling provides better resolution.

Besides the B-mode data used for the 2D image, other information from M-mode and Doppler signal processing can also be displayed. During operation of the ultrasound scanner, information in the memory is continuously updated in real time. When ultrasound scanning is stopped, the last image acquired is displayed on the screen until ultrasound scanning resumes.

Image Storage

Ultrasound images are typically composed of 640×480 or 512×512 pixels. Each pixel typically has a depth of 8 bits (1 byte) of digital data, providing up to 256 lev-

els of gray scale. Image storage (without compression) is approximately $\frac{1}{4}$ megabytes (MB) per image. For real-time imaging (10 to 30 frames per second), this can amount to hundreds of megabytes of data. Color images used for Doppler studies (see section 16.9) increase the storage requirements further because of larger numbers of bits needed for color resolution (full fidelity color requires 24 bits/pixel, one byte each for the red, green, and blue primary colors).

16.7 MISCELLANEOUS ISSUES

Distance, Area, and Volume Measurements

Measurements of distance, area, and volume are performed routinely in diagnostic ultrasound examinations. This is possible because the speed of sound in soft tissue is known to within about $\pm 1\%$ accuracy ($1,540 \text{ m/sec} \pm 15 \text{ m/sec}$), and calibration of the instrument can be easily performed based on round-trip time of the pulse and the echo. A notable example of common distance evaluations are for fetal age determination by measuring fetal head diameter (Fig. 16-38). Measurement accuracy is achieved by careful selection of reference positions, such as the leading edges of two objects along the axis of the beam, as these points are less affected by variations in echo signal amplitude. Measurements between points along the direction of the ultrasound beam are usually more reliable, because of the better spatial resolution in this dimension. Points measured in the lateral plane have a tendency to be blurred in the lateral direction due to the poorer lateral resolution of the system. Thus, horizontal and oblique measurements will be likely to have less accuracy. The circumference of a circular object can easily be calculated from the measured diameter (d) or radius (r), using the relationship circumference of $2\pi r = \pi d$. Distance measurements extend to two dimensions (area) in a straightforward way by assuming a specific geometric shape. Similarly, area measurements in a given image plane extend to 3D volumes by estimating the slice thickness (elevational resolution).

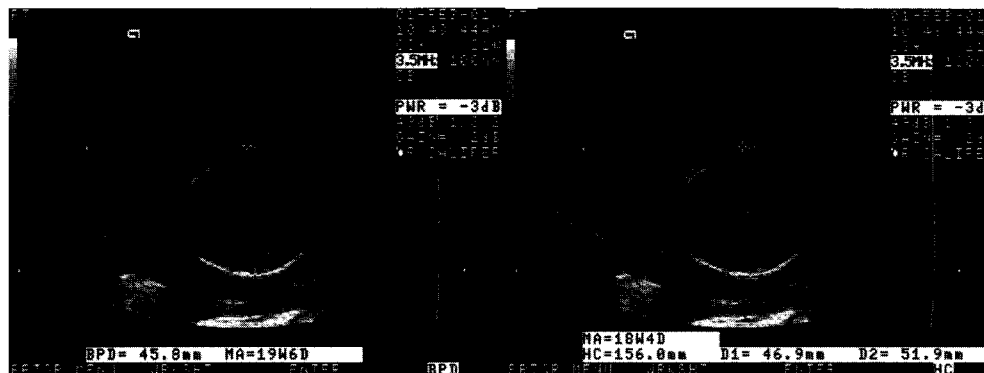


FIGURE 16-38. Ultrasound provides accurate distance measurements. Fetal age is often performed by biparietal (diameter) measurements (*left*) or circumference measurements (*right*). Based on known correlations, the age of the fetus in the images above is estimated to be 19 weeks and 6 days by the diameter assessment, and 18 weeks and 4 days by the circumference measurement.

Ultrasound Contrast Agents

Ultrasound contrast agents for vascular and perfusion imaging are becoming extremely important from the clinical perspective. Most agents are composed of encapsulated microbubbles of 3 to 6 μm diameter containing air, nitrogen, or insoluble gases such as perfluorocarbons. Encapsulation materials, such as human albumin, provide a container for the gas to maintain stability for a reasonable time in the vasculature after injection. The natural tendency is for the gas to diffuse rapidly into the bloodstream (e.g., within seconds) even when encapsulated; successful contrast agents maintain stability over a period of time that allows propagation to specific anatomic vascular areas that are targeted for imaging. Because of the small size of the encapsulated bubbles, perfusion of tissues is also possible, but the bubbles must remain extremely stable during the time required for tissue uptake.

The basis for generating an ultrasound signal is the large difference in acoustic impedance between the gas and the fluids and tissues, as well as the compressibility of the bubbles compared to the incompressible materials that are displaced. The bubbles are small compared with the wavelength of the ultrasound beam and thus become a point source of sound, producing reflections in all directions. In addition, the compressibility produces shifts in the returning frequency of the echoes, called frequency harmonics (described in the next section), which are typically higher than the original ultrasound frequency. To fully use the properties of contrast agents, imaging techniques apart from standard B-mode scans are necessary, and are based on the nonlinear compressibility of the gas bubbles and the frequency harmonics that are generated (see below, e.g., pulse inversion harmonic imaging). Destruction of the microbubbles occurs with the incident ultrasound pulse, and therefore requires temporal delays between images to allow new contrast agent to appear in the FOV.

Harmonic Imaging

Harmonic frequencies are integral multiples of the frequencies contained in an ultrasound pulse; a pulse with a center frequency of f_0 MHz, upon interaction with a medium, will contain high-frequency harmonics of $2f_0$, $3f_0$, $4f_0$, etc. These higher frequencies arise through the vibration of encapsulated gas bubbles used as ultrasound contrast agents or with the nonlinear propagation of the ultrasound as it travels through tissues. For contrast agents, the vibration modes of the encapsulated gas reemit higher-order harmonics due to the small size of the microbubbles (~ 3 to $6 \mu\text{m}$ diameter) and the resultant contraction/expansion from the acoustic pressure variations. Harmonic imaging enhances contrast agent imaging by using a low-frequency incident pulse and tuning the receiver (using a multifrequency transducer) to higher-frequency harmonics. This approach allows removal of “echo clutter” from fundamental frequency reflections in the near field to improve the sensitivity to the ultrasound contrast agent. Even though the returning harmonic signals have higher attenuation (e.g., the first harmonic will have approximately twice the attenuation coefficient compared to the fundamental frequency), the echoes have to only travel half the distance of the originating ultrasound pulse, and thus have a relatively large signal. For harmonic imaging, longer spatial pulse lengths are often used to achieve a higher transducer Q factor. This allows an easier separation of the frequency harmonics from the fundamental frequency. Although the longer SPL degrades axial resolution, the benefits of harmonic imaging overcome the slight degradation in axial resolution.

Based on the harmonic imaging work with microbubble contrast agents, “native tissue” harmonic imaging (imaging higher-frequency harmonics produced by tissues when using lower-frequency incident sound waves) is now possible and in common use. Harmonic frequencies are generated by the nonlinear distortion of the wave as the high-pressure component (compression) travels faster than the low-pressure component (rarefaction) of the acoustic wave in tissue. This wave distortion (Fig. 16-39A), increases with depth, and localizes in the central area of the beam. The returning echoes comprising the harmonics travel only slightly greater than the distance to the transducer, and despite the higher attenuation (due to higher frequency) have less but substantial amplitude compared to the fundamental frequency (Fig. 16-39B). The first harmonic (twice the fundamental frequency) is commonly used because it suffers less attenuation than higher-order harmonics, and because higher-order harmonics are likely to exceed the transducer’s bandwidth. Tuning the broadband receiver to the first harmonic spectrum filters out the lower-frequency echoes (to the extent that the spectra do not overlap) and eliminates ultrasound reflections and scattering from tissues and objects adjacent to the transducer (Fig. 16-39C). Improved lateral spatial resolution (a majority of the echoes are produced in the central area of the beam), reduced side lobe artifacts, and removal of multiple reverberation artifacts caused by anatomy adjacent to the transducer are some advantages of tissue harmonic imaging. Comparison of conven-

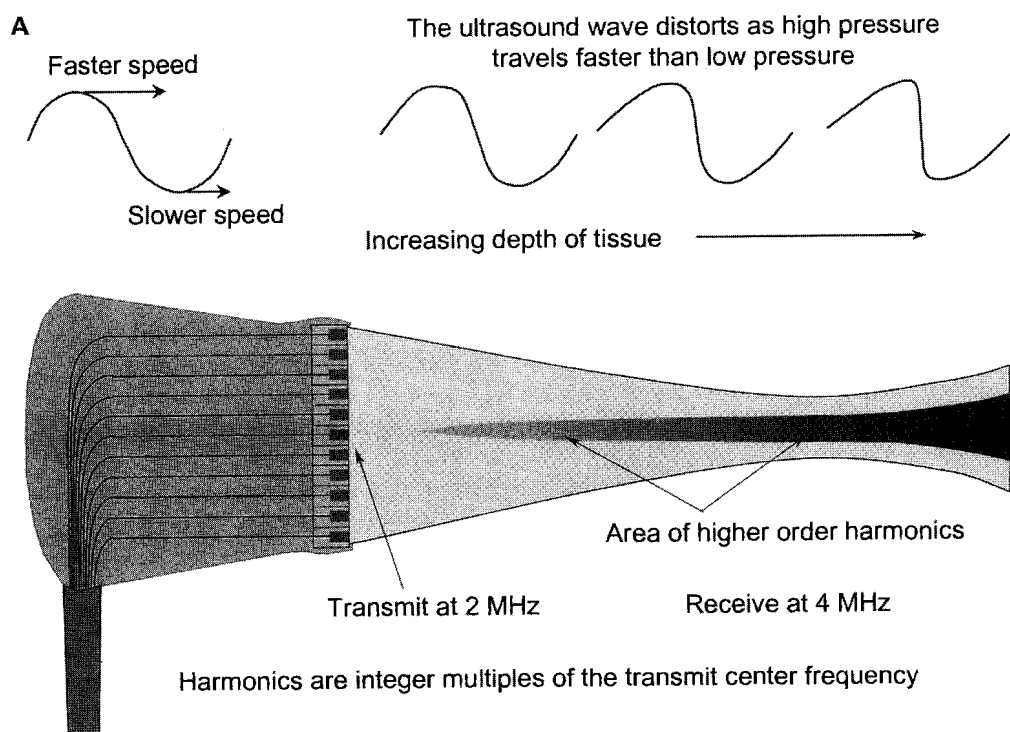


FIGURE 16-39. A: Harmonic frequencies, integer multiples of the fundamental frequency, are produced by the nonlinear propagation of the ultrasound beam, where the high-pressure component travels faster than the low-pressure component of the wave (top). The wave distortion occurs in the central area of the beam.

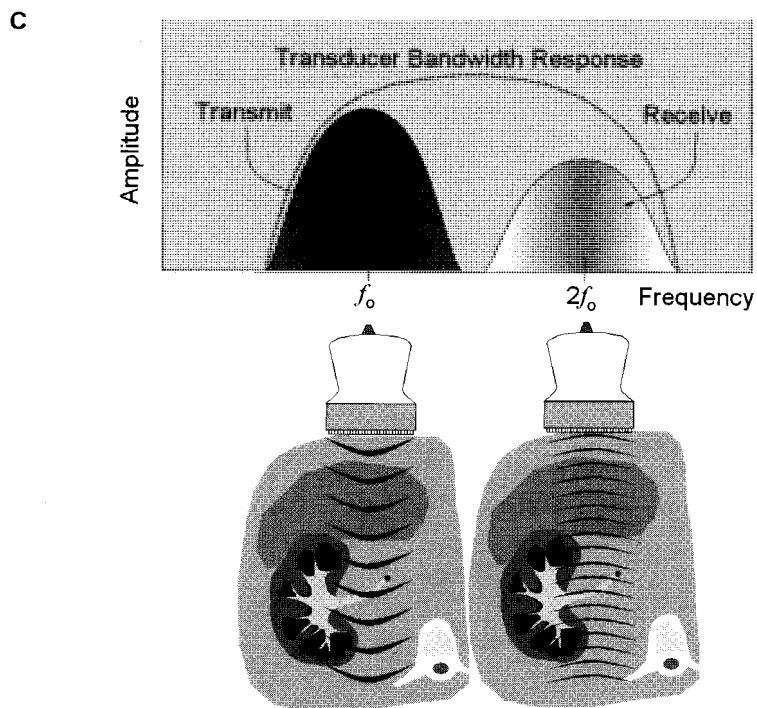
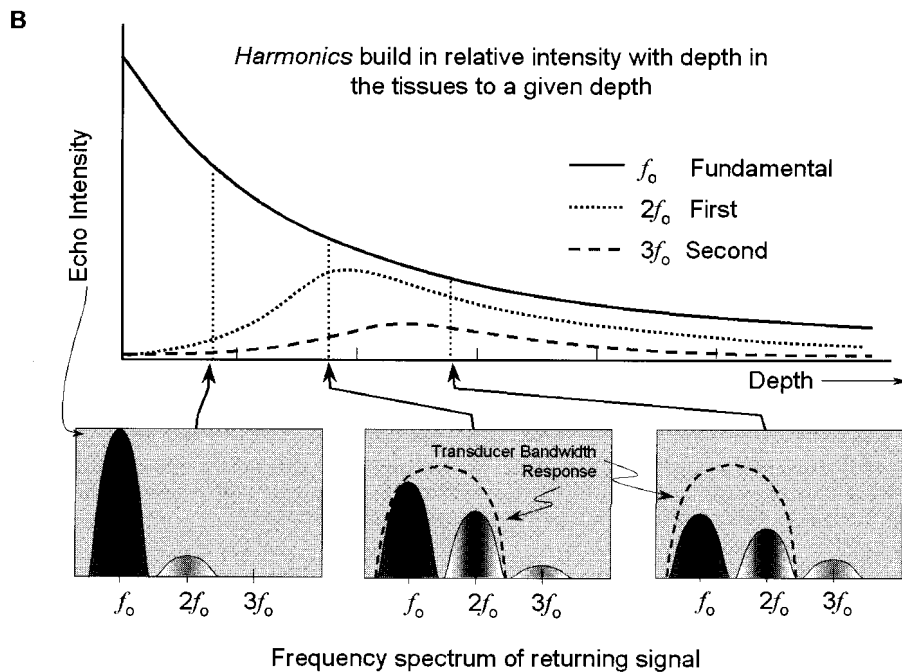


FIGURE 16-39 (continued). **B:** Ultrasound harmonic frequencies ($2f_0$, $3f_0$) increase with depth and attenuate at a higher, frequency-dependent rate. The frequency spectrum and harmonic amplitudes continuously change with depth (*lower figure*, displaying three points in time). The transducer bandwidth response must encompass the higher harmonics. **C:** Native tissue harmonic imaging uses a lower center frequency spectrum (e.g., 2 MHz spectrum) and receives echoes from the higher harmonics (e.g., 4 MHz spectrum). Separation of the transmit and receive frequencies depends on a high Q factor to reduce the incident bandwidth.

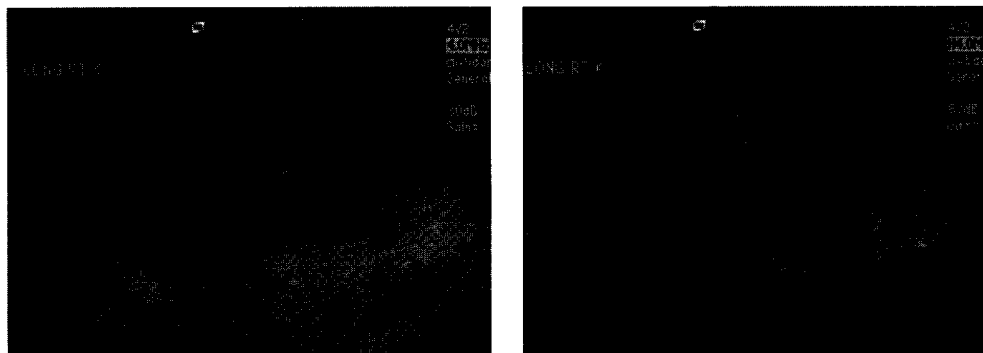


FIGURE 16-40. Conventional (**left**) and harmonic image (**right**) of the kidney demonstrate the improved image quality of the harmonic image from reduced echo clutter and increased contrast. (Courtesy of Kiran Jain, M.D., University of California Davis Health System.)

tional and harmonic right kidney images demonstrates improved image quality (Fig. 16-40). While not always advantageous, native tissue harmonic imaging is best applied in situations such as abdominal imaging that require a lower frequency to achieve adequate penetration to begin with, and then switched to the higher frequency harmonic for better quality and less “clutter” adjacent to the transducer.

A recently introduced imaging technique for ultrasound contrast agents is pulse inversion harmonic imaging, which improves the sensitivity to microbubble contrast agents, and substantially reduces the signals from surrounding soft tissues. Two sequential excitations are acquired for each line in the image. A standard pulse is followed by an inverted (phase-reversed) pulse along the same beam direction. As each beam propagates through the medium, the soft tissues reflect positive and negative pressures back to the transducer with equal but opposite waveforms, and cancel when added. Microbubbles introduced into the vasculature, however, respond nonlinearly (Fig. 16-41A). Harmonic components of the returning echoes are not

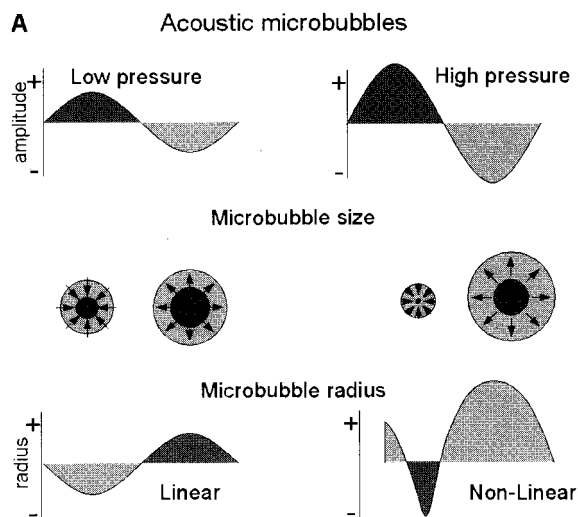


FIGURE 16-41. A: The responses of microbubble contrast agents under low and high pressure reveal nonlinear compression and expansion of the bubble radius (*bottom*).

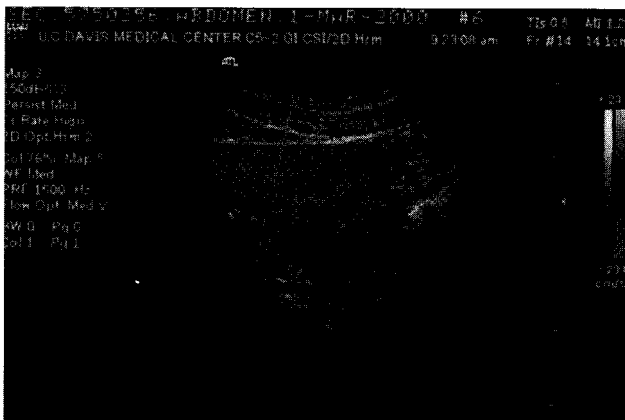
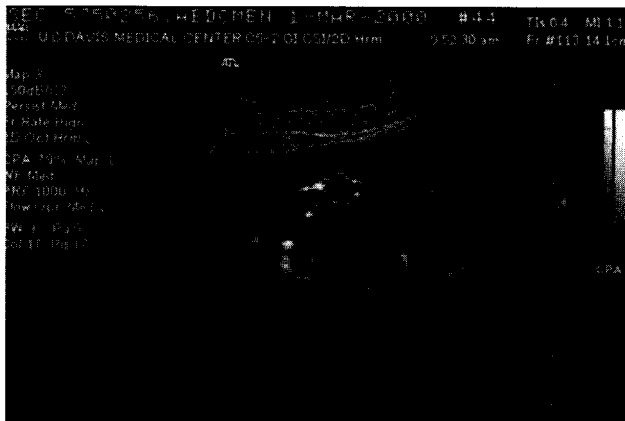
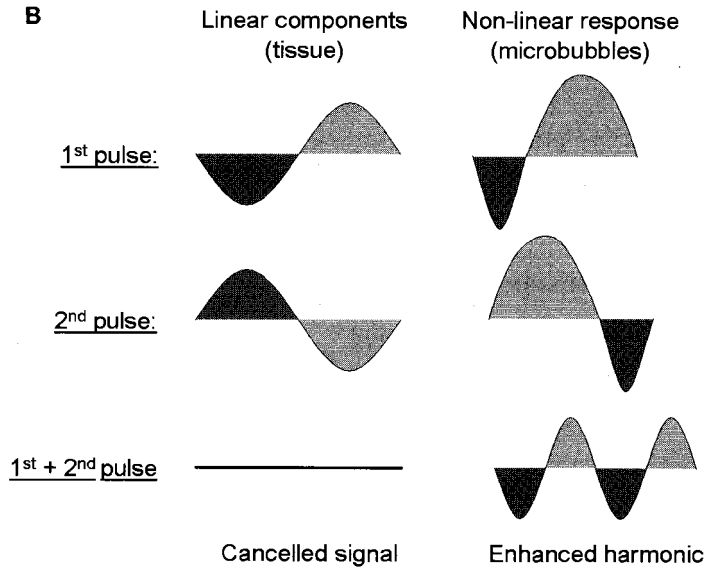


FIGURE 16-41 (continued). B: Pulse-inversion harmonic imaging sends a high-pressure wave, records the echoes, sends a second high-pressure wave of reverse polarity, and adds them together. Linear responses (soft tissue) are canceled, while an enhanced harmonic signal remains, chiefly due to the microbubble contrast agent. **C:** A pulse inversion harmonic ultrasound image shows enhanced sensitivity compared to the conventional color Doppler image (color areas are outlined in the bottom image).

exactly phase reversed, and their sum results in a nonzero (harmonic) signal (Fig. 16-41B). Elimination of competing signals from the surrounding soft tissues (analogous to digital subtraction angiography) enhances the sensitivity to contrast agents, and provides an ability to detect tissue perfusion by microbubble uptake (Fig. 16-41C). Disadvantages include motion artifacts from moving tissues that occur between the pulses, and a frame rate penalty (at least two times slower than a standard scan).

Special Purpose Transducer Assemblies

Most general-purpose ultrasound units are equipped with at least a linear array abdominal transducer of relatively low frequency (good penetration), and a separate small-parts phased array transducer of high frequency (high spatial resolution). Special purpose transducers come in a wide variety of sizes, shapes and operational frequencies, usually designed for a specific application. Intracavitary transducers for transrectal, transurethral, transvaginal, and transesophageal examinations provide high-resolution, high-frequency imaging of anatomic structures that are in proximity to the cavity walls. For example, transducers for transrectal scanning with linear array or mechanical rotating designs are used routinely to image the bladder or the prostate gland. Catheter-mounted rotating single element transducers or phased-array transducers (up to 64 acoustic elements and 20 to 60 MHz operating frequency) create high-resolution (80 to 100 μm detail) acoustic images of vessel lumina. A stack of cylindrical images is created as the catheter is advanced through the vasculature of interest. Intravascular transducers can be used to assess vessel wall morphology, estimate stenosis, and determine the efficacy of vascular intervention. Figure 16-42 shows intravascular ultrasound transducer types, the ultrasound beam, and cylindrical images of a normal and a diseased vessel lumen.

Three-Dimensional Imaging

Three-dimensional ultrasound imaging acquires 2D image data in a series of individual B-scans of a volume of tissue. Forming the 3D data set requires location of each individual 2D image using known acquisition geometries. Volume sampling can be achieved in several ways with a transducer array: (a) linear translation, (b) free-form motion with external localizers to a reference position, (c) rocking motion, and (d) rotation of the transducer (Fig. 16-43 left). With the volume data set acquisition geometry known, rendering into a surface display (Fig. 16-43 right) maximum intensity projection, or multiplanar image reformation is possible using data reordering (see the discussion of 3D imaging in Chapters 13 and 15 for examples and definitions of these techniques). Applications of various 3D imaging protocols are being actively pursued, particularly in obstetrics imaging. In one design, the array transducer is translated perpendicularly to the array direction. This produces a sequence of adjacent sector scans over a 4- to 5-second time period. Each sector scan represents a single plane of the 3D volume according to transducer position. The resultant stack of volume data can be reorganized into image planes to provide alternate tomographic views of the data set. Features such as organ boundaries are identified in each image, and the computer then calculates the 3D surface, complete with shading effects or false color for the delineation of anatomy.

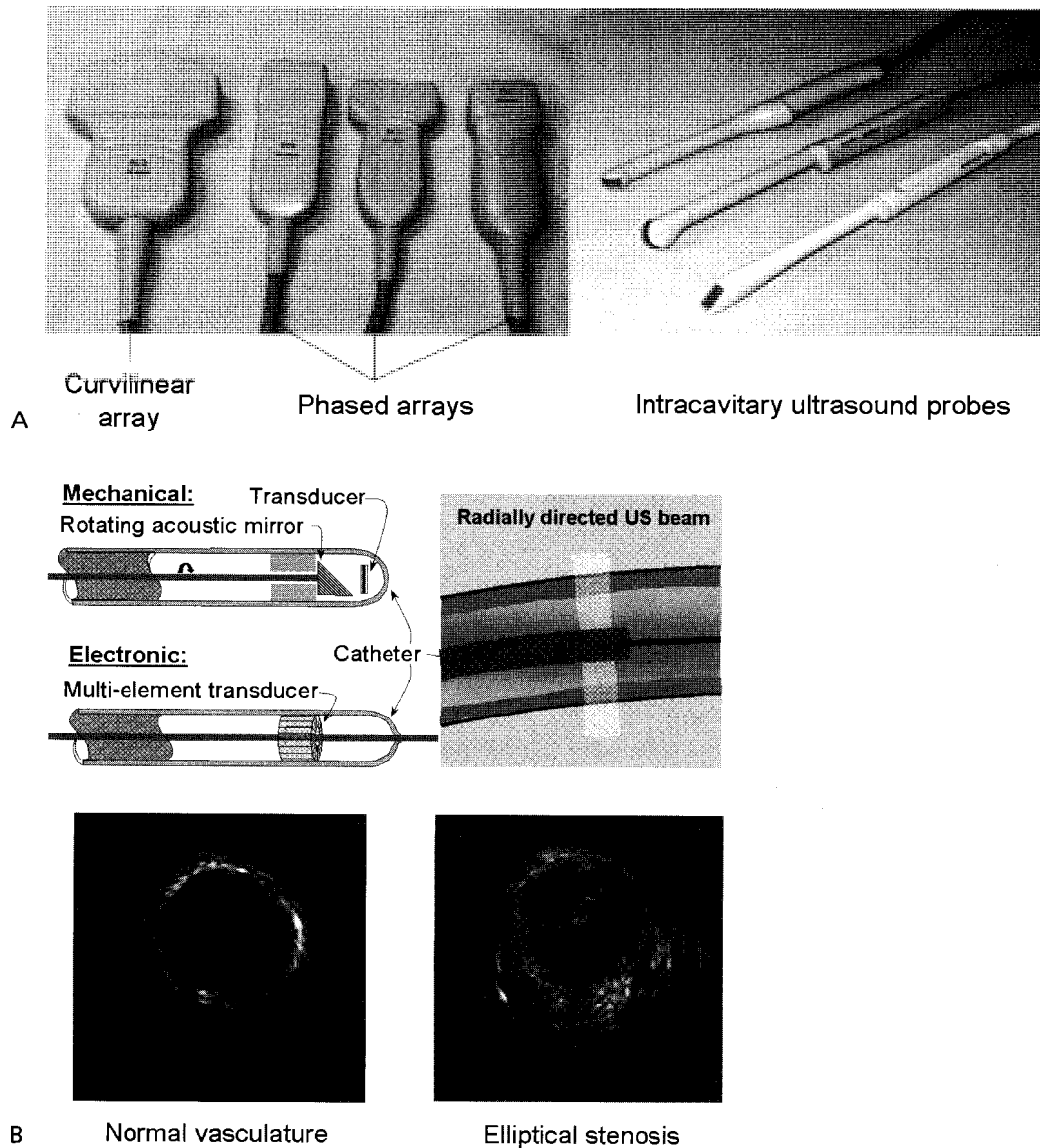


FIGURE 16-42. A: Transducer assemblies come in a variety of shapes and sizes. Conventional (*left*) and intracavitary (*right*) transducers are shown. Intracavitary probes provide an inside-out acoustic mapping of soft tissue anatomy. **B:** Intravascular ultrasound: devices and images. Catheter-mounted transducer arrays provide an acoustic analysis of the vessel lumen and wall from the inside out. Mechanical (rotating shaft and acoustic mirror with a single transducer) and electronic phased array transducer assemblies are used. Images show a normal vessel lumen (*lower left*) and reduced lumen due to stenosis (*lower right*).

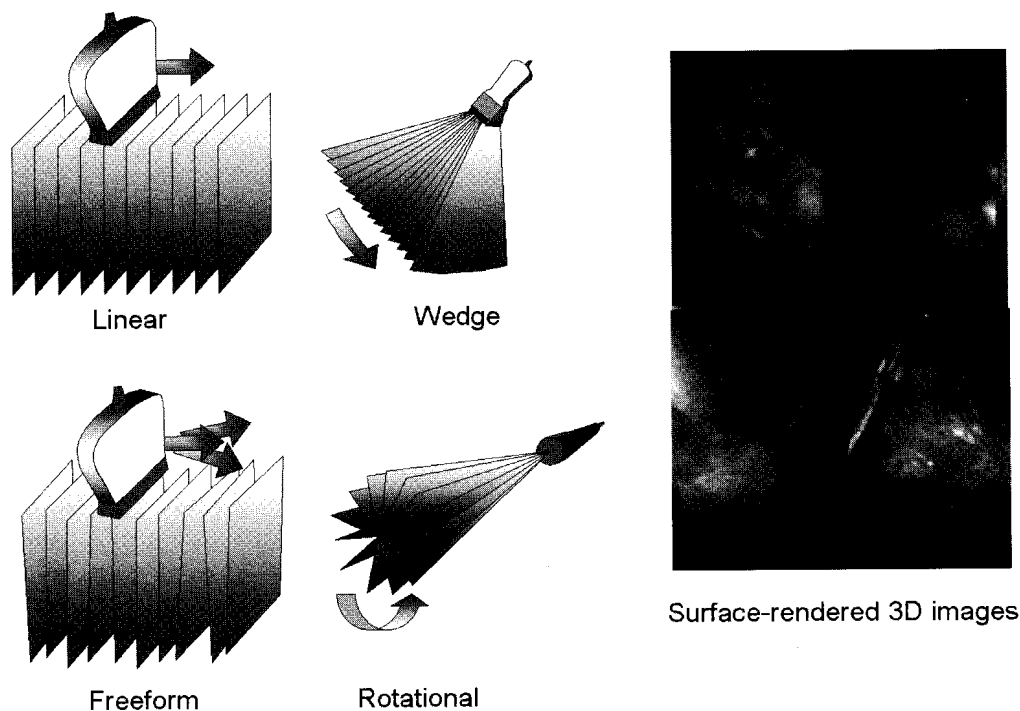


FIGURE 16-43. Three-dimensional ultrasound acquisitions can be accomplished in several ways (left), and include linear, wedge, freeform, and circular formats. Reconstruction of the data set provides 3D surface-shaded and/or wire mesh renditions of the anatomy. Depicted is a 3D surface display image of a fetus *in situ*, of normal development (**top**) and a cleft palate (**bottom**). (Courtesy of Dr. Thomas Nelson and Dr. Delores Pretorius, of the University of California San Diego.)

16.8 IMAGE QUALITY AND ARTIFACTS

Image quality is dependent on the design characteristics of the ultrasound equipment, the numerous equipment variables selected, and positioning skills of the operator. The equipment variables controlled by the operator include transducer frequency, pulse repetition frequency, ultrasound intensity, and TGC curves, among others. Measures of ultrasound image quality include spatial resolution, contrast resolution, image uniformity, and noise characteristics. Image artifacts are common phenomena that can enhance or degrade the diagnostic value of the ultrasound image.

Spatial Resolution

Ultrasound spatial resolution has components in three directions—axial, lateral, and elevational. Axial resolution is determined by the frequency of the ultrasound and the damping factor of the transducer, which together determine the spatial pulse length. Resolution in the axial direction (along the beam) is equal to $\frac{1}{2}$ SPL, and independent of depth. Lateral and elevational resolutions are strongly dependent on depth. Lateral and elevational resolutions are determined by the dimensions

(width and height, respectively) of the transducer aperture, the depth of the object, and mechanical and electronic focusing. Axial and lateral resolutions are in the plane of the image and plainly discernible, while elevational (slice thickness) resolution is perpendicular to the plane of the image, and not as easy to understand or to interpret. The minimum resolution in the lateral/elevational directions is typically 3 to 5 times worse than axial resolution.

Elevational resolution is a function of the height of the transducer array and is depth dependent. Poor elevational resolution occurs adjacent to the transducer array, and beyond the near/far field interface. Elevational focusing is possible with an acoustic lens shaped along the height of the elements, which can produce an elevational focal zone closer to the array surface. Alternatively, "1.5-D" array transducers have several rows (typically five to seven) of independent elements in the elevational direction ("1.5-D" indicates that the number of elements in the elevational direction is much less than in the lateral direction). Elevational focusing is achieved by introducing phase delays among the elements in different rows to electronically focus the beam in the slice thickness dimension at a given depth (Fig. 16-25). Multiple elevational transmit focal zones incur a time penalty similar to that required for the multiple lateral focal zones.

Contrast Resolution and Noise

Contrast resolution depends on several interrelated factors. Acoustic impedance differences (Table 16-4) give rise to reflections that delineate tissue boundaries and internal architectures. The density and size of scatterers within tissues or organs produce a specific "texture" (or lack of texture) that permits recognition and detection over the field of view. With proper signal processing, attenuation differences (Table 16-5) result in gray-scale differences among the tissues. Areas of low and high attenuation often produce distal signal enhancement or signal loss (e.g., fluid-filled cysts, gallstones) that allows detection and identification of tissues in the image. Introduction of microbubble contrast agents improves the visualization of the vasculature and permits the assessment of tissue perfusion. Harmonic imaging improves image contrast by eliminating unimportant or degrading signals from lower-frequency echoes. Pulse inversion harmonic imaging is a new technique that enhances contrast resolution further. In addition, Doppler imaging techniques (discussed in the next section of this chapter) uses moving anatomy (blood flow) and sophisticated processing techniques to generate contrast.

Contrast resolution also depends on spatial resolution. Details within the image are often distributed over the volume element represented in the tomographic slice (e.g., Fig. 16-21), which varies in size in the lateral and elevational dimensions as a function of depth. In areas where the slice thickness is relatively wide (close to the transducer array surface and at great depth), the returning echoes generated by a small object will be averaged over the volume element, resulting in a lower signal and loss of detection (see Fig. 16-55B). On the other hand, objects larger than the minimum volume element can actually achieve better contrast relative to the background because the random noise components are reduced by averaging over the volume.

Detection of subtle anatomy in the patient is dependent on the contrast-to-noise ratio. The contrast is generated by differences in signal amplitude as discussed above. Noise is mainly generated by the electronic amplifiers of the system but is

occasionally induced by environmental sources such as electrical power fluctuations and equipment malfunction such as a dead or poorly functioning transducer element. A low-noise, high-gain amplifier is crucial for optimal low-contrast resolution. Exponential attenuation of the ultrasound beam requires time gain compensation, which reduces contrast and increases noise with depth. Image processing that specifically reduces noise, such as temporal or spatial averaging, can increase the contrast-to-noise ratio; however, lower frame rates and/or poorer spatial resolution are trade-offs. Low-power operation (e.g., an obstetrics power setting with low transmit gain) requires higher electronic signal amplification to increase the weak echo amplitudes to useful levels, and results in a lower contrast-to-noise ratio. Increasing the transmit power and/or the pulse repetition frequency can improve contrast resolution, but there is a limit with respect to transducer capabilities, and, furthermore, the intensity must be restricted to levels unlikely to cause biologic damage.

Artifacts

Artifacts arise from the incorrect display of anatomy or noise during imaging. The causes are machine and operator related, as well as intrinsic to the characteristics and interaction of ultrasound with the tissues. Understanding how artifacts are generated in ultrasound is crucial, which places high demands on the knowledge and experience of the sonographer.

Artifacts can be caused by a variety of mechanisms. For instance, sound travels at different speeds, not just the 1,540 m/sec average value for soft tissue that is used in the “range equation” for placing echo information in the image matrix. This speed variation results in some echoes being displaced from the expected location in the image display. Sound is refracted when the beam is not perpendicular to a tissue boundary; echoes are deflected to the receiver from areas outside of the main beam and mismapped into the image. Improper use of TGC causes suboptimal image quality with nonuniform appearance of tissues within a given band of the image.

Fortunately, most ultrasound artifacts are readily discernible because of transient appearance and/or obvious effects on the image. Some artifacts are used to advantage as a diagnostic aid in characterization of tissue structures and their composition. Many common artifacts are discussed below.

Refraction

Refraction is a change in the transmitted ultrasound pulse direction at a boundary with nonperpendicular incidence, when the two tissues have different speeds of sound ($c_1 \neq c_2$), causing misplaced anatomic position in the image (Fig. 16-44A). Anatomic displacement due to refraction artifacts will change with the position of the transducer and angle of incidence with the tissue boundaries.

Shadowing and Enhancement

Shadowing is a hypointense signal area distal to an object or interface, and is caused by objects with high attenuation or reflection of the incident beam. Highly attenuating objects such as bones or kidney stones reduce the intensity of the

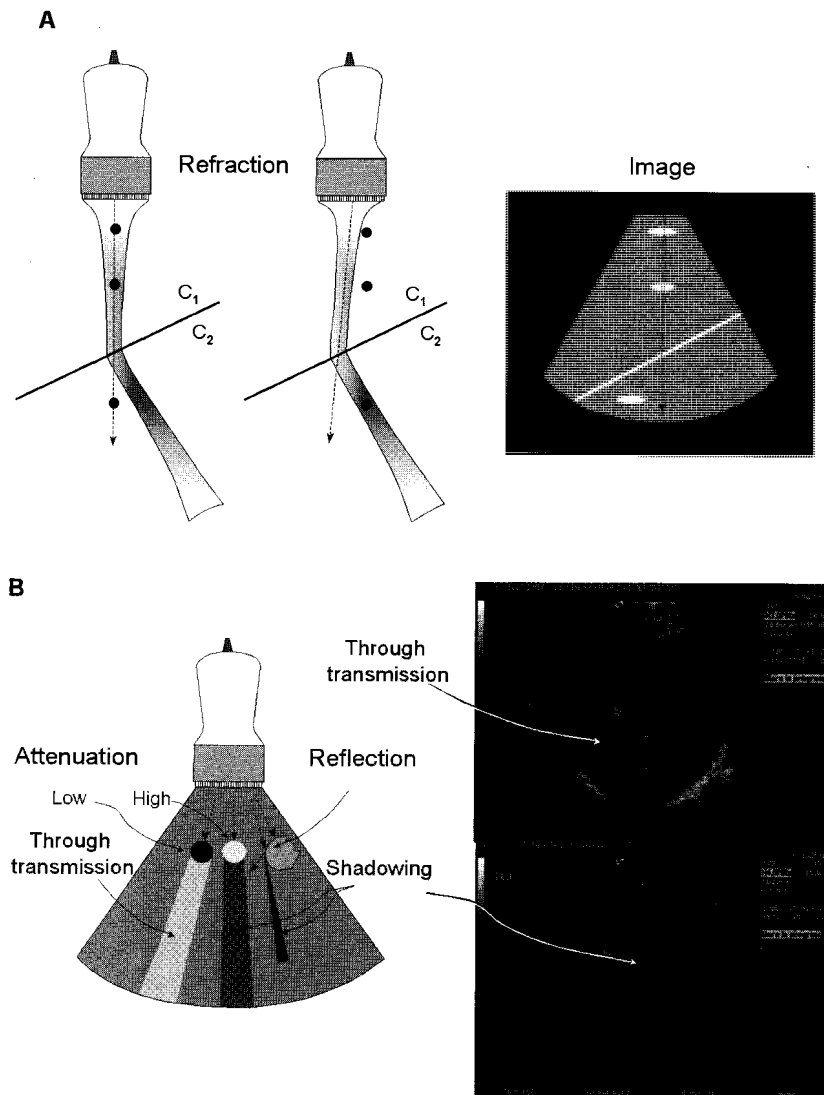


FIGURE 16-44. Common ultrasound artifacts. **A:** Refraction is a change in the direction of the ultrasound beam that results from nonperpendicular incidence at a boundary where the two tissues do not have the same speed of sound. During a scan, anatomy can be missed and/or mislocated from the true position. **B:** Attenuation and reflection of the ultrasound beam causes intensity changes. Enhancement occurs distal to objects of low attenuation, manifested by a hyperintense signal, while shadowing occurs distal to objects of high attenuation, resulting in a hypointense signal. At the curved edges of an ultrasound mass, nonperpendicular reflection can cause distal hypointense streaks beyond the mass. Clinical images show through transmission (top) of a low attenuation cyst, and shadowing (bottom), caused by high attenuation gallstones.

(continued on next page)

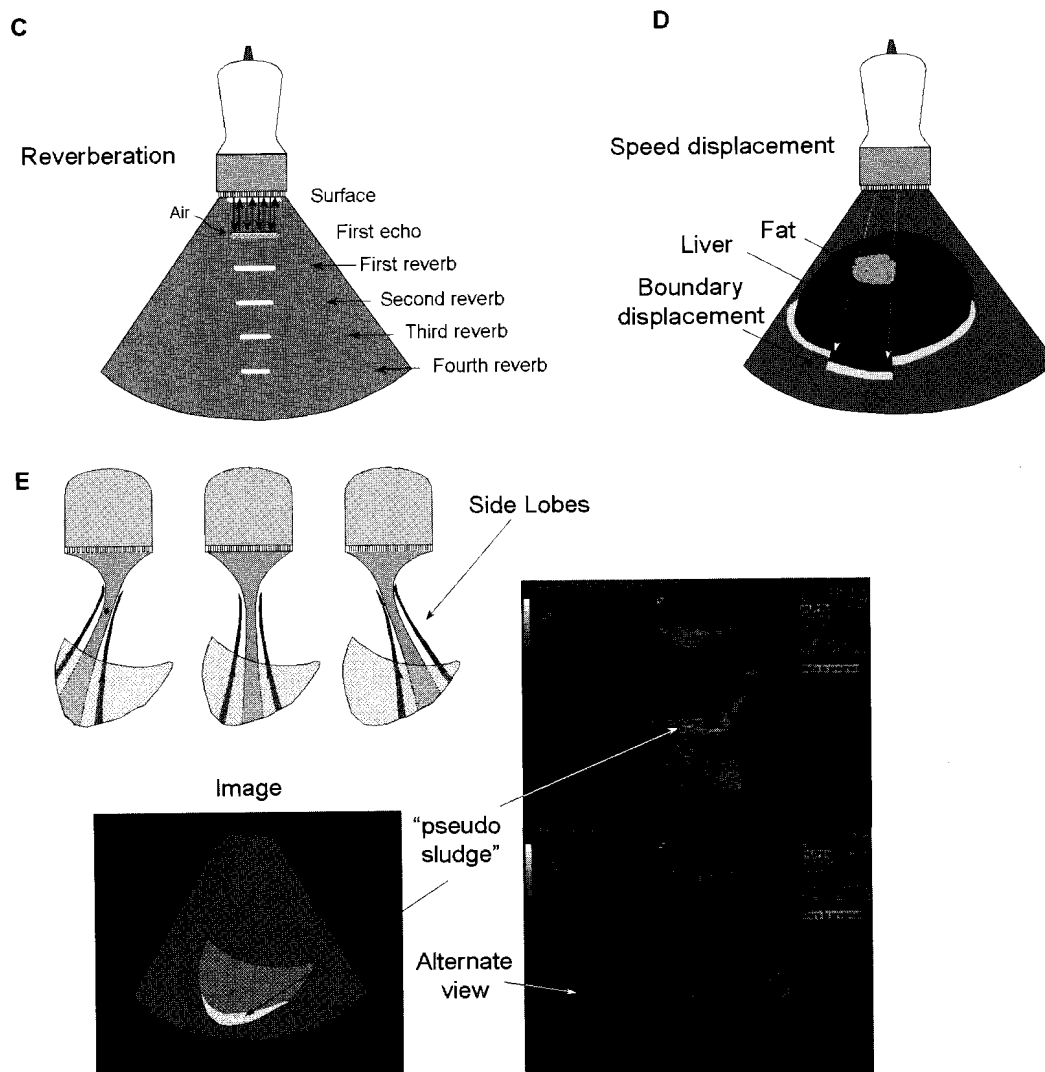


FIGURE 16-44 (continued). **C:** Reverberation commonly occurs between two strong reflectors, such as an air pocket and the transducer array at the skin surface. The echoes bounce back and forth between the two boundaries and produce equally spaced signals of diminishing amplitude in the image. This is often called a "comet-tail" artifact. **D:** Speed of sound variation in the tissues can cause a mismatching of anatomy. In the case of fatty tissues, the slower speed of sound in fat (1,450 m/sec) results in a displacement of the returning echoes from distal anatomy by about 6% of the distance traveled through the mass. **E:** Side-lobe energy emissions in transducer arrays can cause anatomy outside of the main beam to be mapped into the main beam. For a curved boundary, such as the gallbladder, side-lobe interactions can be remapped and produce findings such as "pseudo-sludge" that is not apparent with other scanning angles. Clinical images (*top*) show pseudo-sludge, which is not evident after repositioning the transducer assembly (*bottom*).

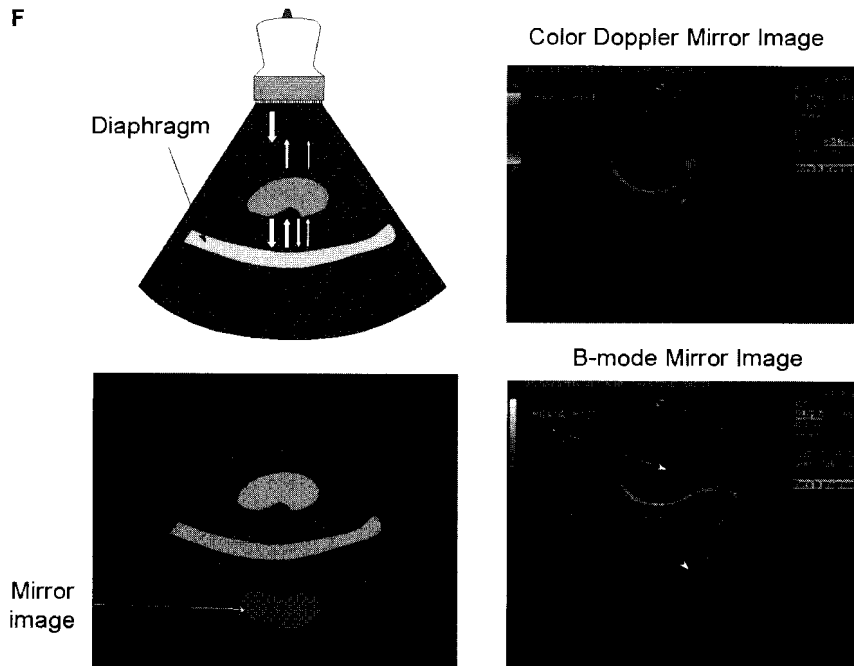


FIGURE 16-44 (continued). F: A mirror image artifact arises from multiple beam reflections between a mass and a strong reflector, such as the diaphragm. Multiple echoes result in the creation of a mirror image beyond the diaphragm of the mass. Clinical images show mirror artifacts for color Doppler (*top*) and B-mode scan (*bottom*), where the *lower arrow* points to the mirror artifactual appearance of the diaphragm.

transmitted beam, and can induce low-intensity streaks in the image. Reflection of the incident beam from curved surfaces eliminates the distal propagation of ultrasound and causes streaks. Enhancement occurs distal to objects having very low ultrasound attenuation, such as fluid-filled cavities (e.g., a filled bladder or cysts). Hyperintense signals arise from increased transmission of sound by these structures (Fig. 16-44B).

Reverberation

Reverberation artifacts arise from multiple echoes generated between two closely spaced interfaces reflecting ultrasound energy back and forth during the acquisition of the signal and before the next pulse. These artifacts are often caused by reflections between a highly reflective interface and the transducer, or between reflective interfaces such as metallic objects (e.g., bullet fragments) or air pocket/partial liquid areas of the anatomy. Reverberation echoes are typically manifested as multiple, equally spaced boundaries with decreasing amplitude along a straight line from the transducer (Fig. 16-44C). They are also called ring-down or comet tail artifacts.

quency of the reflected sound. Thus, the Doppler shift is nearly proportional to the

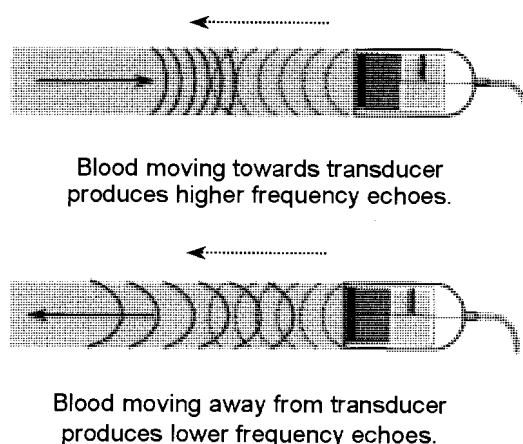


FIGURE 16-45. Doppler ultrasound. Sound waves reflected from a moving object are compressed (higher frequency) when moving toward the transducer, and expanded (lower frequency) when moving away from the transducer compared to the incident sound wave frequency. The difference between the incident and returning frequencies is called the Doppler shift frequency.

velocity of the blood cells. When the sound waves and blood cells are not moving in parallel directions, the equation must be modified to account for less Doppler shift. The angle between the direction of blood flow and the direction of the sound is called the Doppler angle (Fig. 16-46). The component of the velocity vector directed toward the transducer is the velocity along the vessel axis multiplied by the cosine of the angle, $\cos(\theta)$. Without correction for this angular dependence, the smaller Doppler shift will cause an underestimate of the actual blood velocity. The Doppler angle joins the adjacent side and hypotenuse of a right triangle; therefore, the component of the blood velocity in the direction of the sound (adjacent side) is equal to the actual blood velocity (hypotenuse) multiplied by the cosine of the angle (θ). [Note: In physics, the term *velocity* refers to a vector quantity, describing both the amount of distance traveled per unit time (speed) and the direction of movement (such as blood flow). However, it is common practice in medical ultrasound to use the term *velocity* instead of the proper term *speed*.] As the velocity of blood cells (peak of ~ 200 cm/sec) is significantly less than the speed of sound (154,000 cm/sec), the denominator can be simplified with extremely small

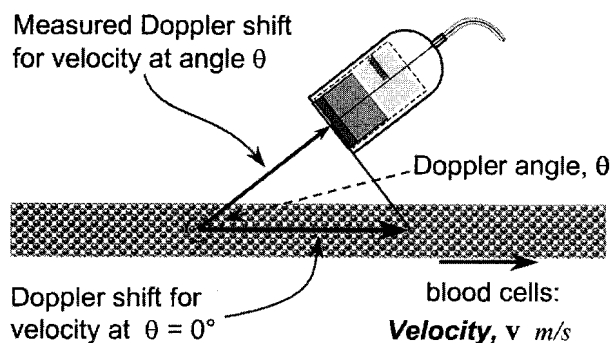


FIGURE 16-46. Geometry of Doppler ultrasound data acquisition. The Doppler shift varies as a function of angle (θ) of the incident ultrasound pulse and the axis of the blood vessel for a fixed blood velocity v . The maximum Doppler shift occurs at an angle $\theta = 0$. For a larger Doppler angle (θ), the measured Doppler shift is less by a factor of $\cos(\theta)$, and velocity estimates are compensated by $1/\cos(\theta)$.

error by neglecting the velocity of the blood. This results in the generalized Doppler shift equation:

$$f_d = \frac{2 f_i v \cos(\theta)}{c}$$

where v is the velocity of blood, c is the speed of sound in soft tissue, and θ is the Doppler angle.

The blood velocity can be calculated by rearranging the Doppler equation:

$$v = \frac{f_d c}{2 f_i \cos(\theta)}$$

Thus, the measured Doppler shift at a Doppler angle θ is adjusted by $1/\cos(\theta)$ to achieve accurate velocity estimates. Selected cosine values are $\cos 0$ degrees = 1, $\cos 30$ degrees = 0.87, $\cos 45$ degrees = 0.707, $\cos 60$ degrees = 0.5, and $\cos 90$ degrees = 0. At a 60-degree Doppler angle, the measured Doppler frequency shift is one-half the actual Doppler frequency, and at 90 degrees the measured frequency shift is 0. The preferred Doppler angle ranges from 30 to 60 degrees. At too large an angle (>60 degrees), the apparent Doppler shift is small, and minor errors in angle accuracy can result in large errors in velocity (Table 16-7). At too small an angle (e.g., <20 degrees), refraction and critical angle interactions can cause problems, as can aliasing of the signal in pulsed Doppler studies.

The Doppler frequency shifts for moving blood occur in the audible range. It is both customary and convenient to convert these frequency shifts into an audible signal through a loudspeaker that can be heard by the sonographer to aid in positioning and to assist in diagnosis.

Example: Given: $f_i = 5$ MHz, $v = 35$ cm/sec, and $\theta = 45$ degrees, calculate the Doppler frequency shift.

$$f_d = \frac{2 \times 5 \times 10^6/\text{sec} \times 35 \text{ cm/sec} \times 0.707}{154,000 \text{ cm/sec}} = 1.6 \times 10^3/\text{sec} = 1.6 \text{ kHz}$$

The frequency shift of 1.6 kHz is in the audible range (15 Hz to 20 kHz). With an increased Doppler angle, the measured Doppler shift is decreased according to $\cos(\theta)$ since the projection of the velocity vector toward the transducer decreases (and thus the audible frequencies decrease to lower pitch).

TABLE 16-7. DOPPLER ANGLE AND ERROR ESTIMATES OF BLOOD VELOCITY FOR A +3-DEGREE ANGLE ACCURACY ERROR

Angle (degrees)	Set Angle (degree)	Actual Velocity (cm/sec)	Estimated Velocity (cm/sec)	Error (%)
0	3°	100	100.1	0.14
25	28°	100	102.6	2.65
45	48°	100	105.7	5.68
60	63°	100	110.1	10.1
80	83°	100	142.5	42.5

Continuous Doppler Operation

The continuous-wave Doppler system is the simplest and least expensive device for measuring blood velocity. Two transducers are required, with one transmitting the incident ultrasound and the other detecting the resultant continuous echoes (Fig. 16-47). An oscillator produces a resonant frequency to drive the transmit transducer and provides the same frequency signal to the demodulator, which compares the returning frequency to the incident frequency. The receiver amplifies the returning signal, and the mixer demodulator extracts the Doppler shift frequency by using a "low-pass" filter, which removes the superimposed high-frequency oscillations. The Doppler signal contains very low frequency signals from vessel walls and other moving specular reflectors that a "wall filter" selectively removes. An audio amplifier amplifies the Doppler signal to an audible sound level, and a recorder tracks spectrum changes as a function of time for analysis of transient pulsatile flow.

Continuous-wave Doppler suffers from depth selectivity with accuracy affected by object motion within the beam path. Multiple overlying vessels will result in superimposition, making it difficult to distinguish a specific Doppler signal. Spectral broadening of the frequencies occurs with a large sample area across the vessel profile (composed of high velocity in the center and slower velocities at the edge of the vessels). Advantages of continuous mode include high accuracy of the Doppler

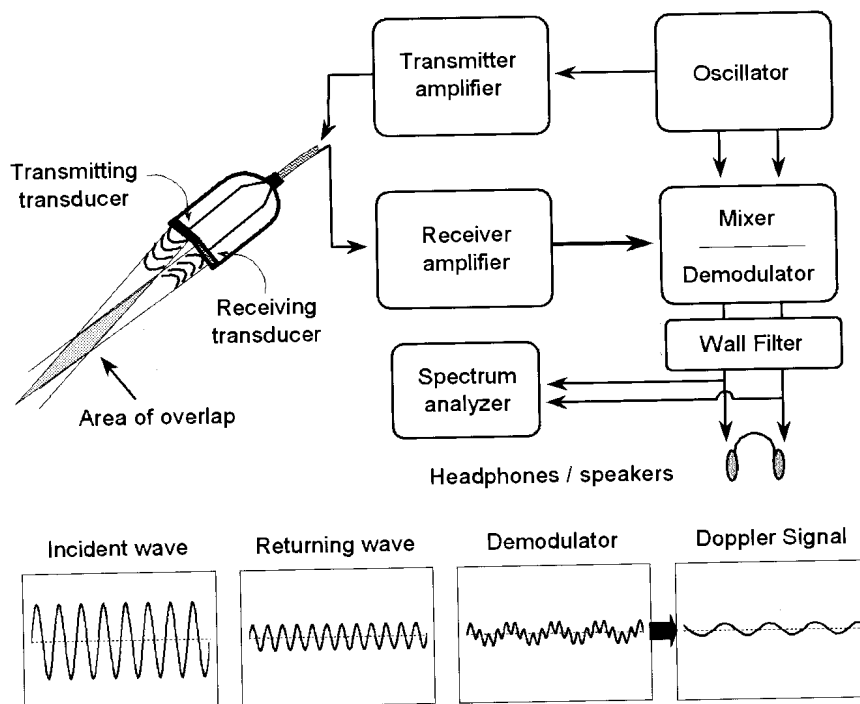


FIGURE 16-47. Block diagram of a continuous-wave Doppler system. Two transducers are required: one as a transmitter and the other as a receiver. The area of overlap determines the position of blood velocity measurement. Signals from the receiver are mixed with the original frequency to extract the Doppler signal. A low-pass filter removes the highest frequencies in the demodulated signals, and a high-pass filter (Wall filter) removes the lowest frequencies due to tissue and transducer motion to extract the desired Doppler shift.

shift measurement because a narrow frequency bandwidth is used, and that high velocities are measured without aliasing (see Pulsed Doppler Operation, below).

Quadrature Detection

The demodulation technique measures the magnitude of the Doppler shift, but does not reveal the direction of the Doppler shift, i.e., whether the flow is toward or away from the transducers. A method of signal processing called quadrature detection, beyond the scope of this discussion, permits determination of whether the direction of flow is toward or away from the transducers.

Pulsed Doppler Operation

Pulsed Doppler ultrasound combines the velocity determination of continuous-wave Doppler systems and the range discrimination of pulse echo imaging. A transducer tuned for pulsed Doppler operation is used in a pulse echo format, similar to imaging. The spatial pulse length is longer (a minimum of 5 cycles per pulse up to 25 cycles per pulse) to improve the measurement accuracy of the frequency shift (although at the expense of axial resolution). Depth selection is achieved with an electronic time gate circuit to reject all echo signals except those falling within the gate window, as determined by the operator. In some systems, multiple gates provide profile patterns of velocity values across a vessel. Figure 16-48 is a simple diagram of a pulsed Doppler system.

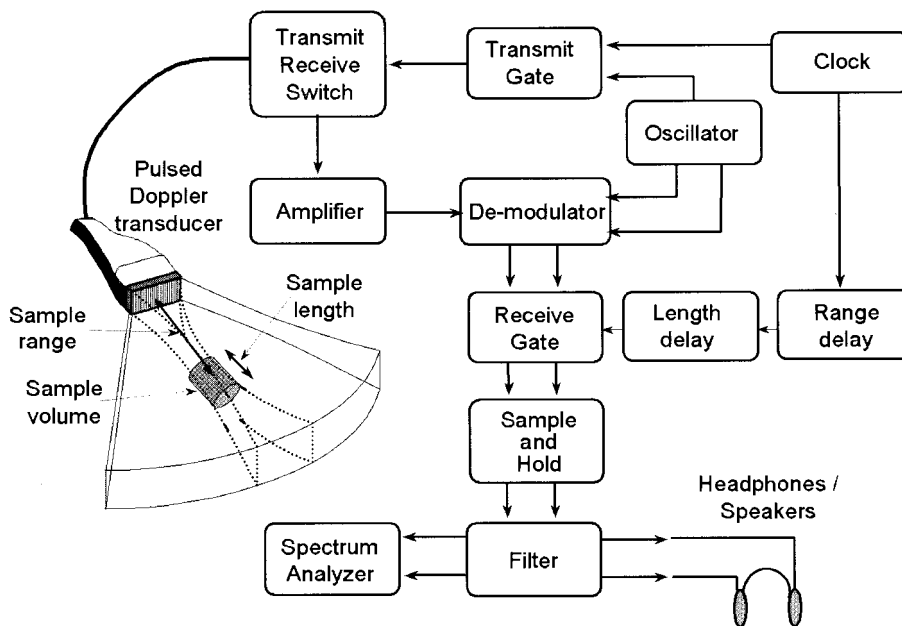


FIGURE 16-48. Block diagram of a pulsed Doppler system. Isolation of a selected area is achieved by gating the time of echo return, and analyzing only those echoes that fall within the time window of the gate. In the pulsed mode, the Doppler signal is discretely sampled in time to estimate the frequency shifts occurring in the Doppler gate. Because axial resolution isn't as important as narrow bandwidths to better estimate the Doppler shift, a long spatial pulse width (high Q factor) is employed.

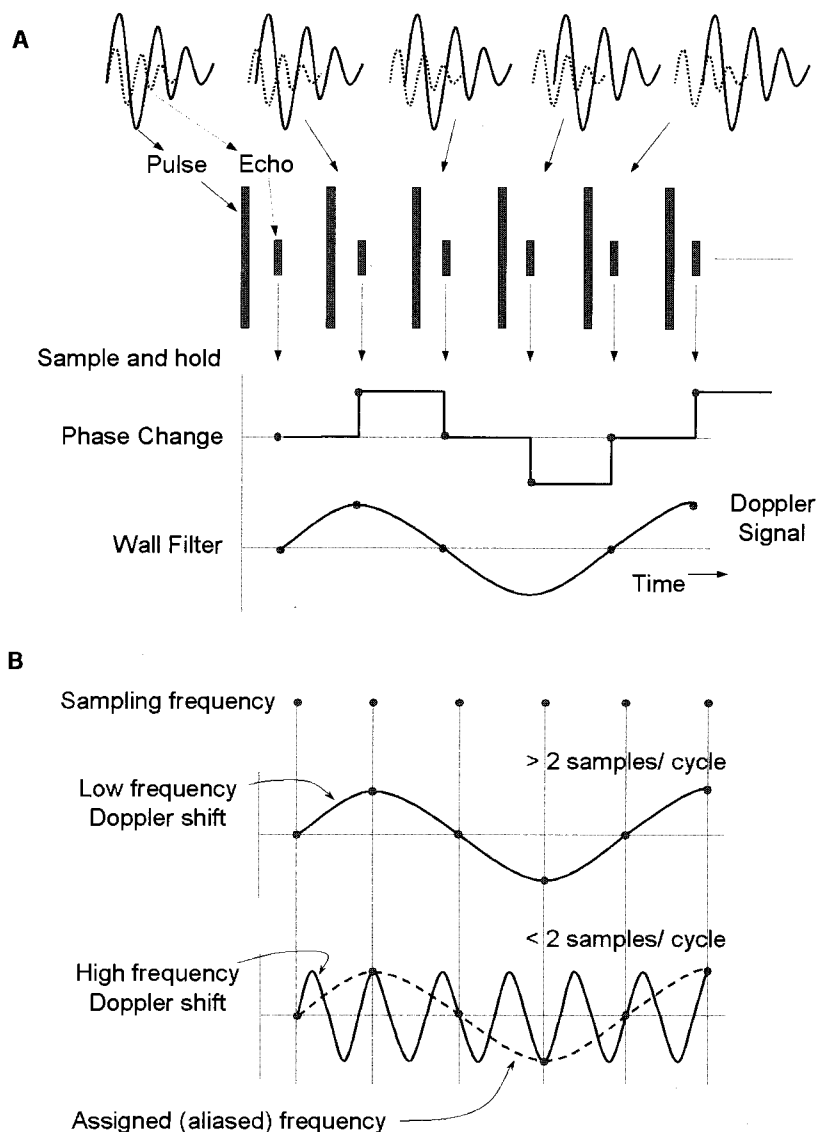


FIGURE 16-49. A: The returning ultrasound pulses from the Doppler gate are sampled over several pulse echo cycles (in this example, five times), to estimate Doppler shifts (if any) caused by moving blood cells. A sample and hold circuit produces a variation of signal with time [note the echo (*dashed line*) and the pulse (*solid line*) vary in phase], which are recorded and analyzed by Fourier transform methods to determine the Doppler shift frequencies. A wall filter removes the low-frequency degradations caused by transducer and patient motion. **B:** Aliasing occurs when the frequencies in the sampled signal are greater than one-half the PRF (sampling frequency). In this example, a signal of twice the frequency is analyzed as if it were the lower frequency, and thus “masquerades” as the lower frequency.

Each Doppler pulse does not contain enough information to completely determine the Doppler shift, but only a sample of the shifted frequencies measured as a phase change. The phases of the returning echoes from a stationary object, relative to the phase of the transmitted sound, does not change with time. However, the phases of the returning echoes from moving objects do vary with time. Repeated echoes from the active gate are analyzed in the sample/hold circuit, and a Doppler signal is gradually built up (Fig. 16-49A). The discrete measurements acquired at the pulse repetition frequency (PRF) produce the synthesized signal. According to sampling theory, a signal can be reconstructed unambiguously as long as the true frequency (e.g., the Doppler shift) is less than half the sampling rate. Thus, the PRF must be at least twice the maximal Doppler frequency shift encountered in the measurement.

The maximum Doppler shift $\Delta f_{\max}$ that is unambiguously determined in the pulsed Doppler acquisition follows directly from the Doppler equation by substituting $V_{\max}$ for V :

$$\Delta f_{\max} = \frac{\text{PRF}}{2} = \frac{2f_0 V_{\max} \cos(\theta)}{c}$$

Rearranging the equation and solving for $V_{\max}$,

$$V_{\max} = \frac{c \times \text{PRF}}{4 \times f_0 \times \cos(\theta)}$$

shows that the maximum blood velocity that is accurately determined is increased with larger PRF, lower operating frequency, and larger angle (the cosine of the angle gets smaller with larger angles from 0 to 90 degrees).

For Doppler shift frequencies exceeding one-half the PRF, aliasing will occur, causing a potentially significant error in the velocity estimation of the blood (Fig. 16-49B). A 1.6-kHz Doppler shift requires a minimum PRF of $2 \times 1.6 \text{ kHz} = 3.2 \text{ kHz}$. One cannot simply increase the PRF to arbitrarily high values, because of echo transit time and possible echo ambiguity. Use of a larger angle between the ultrasound beam direction and the blood flow direction (e.g., 60 degrees) reduces the Doppler shift. Thus, higher velocities can be unambiguously determined for a given PRF at larger angles. At larger angles (e.g., 60 to 90 degrees), however, small errors in angle estimation cause significant errors in the estimation of blood velocity (Table 16-7).

A 180-degree phase shift in the Doppler frequency represents blood that is moving away from the transducer. Often, higher frequency signals in the spectrum will be interpreted as lower frequency signals with a 180-degree phase shift, such that the highest blood velocities in the center of a vessel are measured as having reverse flow. This is another example of aliasing.

Duplex Scanning

Duplex scanning refers to the combination of 2D B-mode imaging and pulsed Doppler data acquisition. Without visual guidance to the vessel of interest, pulsed Doppler systems would be of little use. A duplex scanner typically operates in the 2D B-mode to create the real-time image to facilitate selecting the Doppler gate window position, and then is switched to the Doppler mode where the ultrasound

beam is aligned at a particular orientation and with appropriate range delay to obtain Doppler information.

Instrumentation for duplex scanning is available in several configurations. Most often, electronic array transducers switch between a group of transducers used to create a B-mode image and one or more transducers used for the Doppler information.

The duplex system allows estimation of the flow velocity directly from the Doppler shift frequency, since the velocity of sound and the transducer frequency are known, while the Doppler angle can be estimated from the B-mode image by the user and input into the scanner computer for calculation. Once the velocity is known, flow (in units of cm^3/sec) is estimated as the product of the vessel's cross-sectional area (cm^2) times the velocity (cm/sec).

Errors in the flow volume may occur. The vessel axis might not lie totally within the scanned plane, the vessel might be curved, or flow might be altered from the perceived direction. The beam-vessel angle (Doppler angle) could be in error, which is much more problematic for very large angles, particularly those greater than 60 degrees, as explained previously. The Doppler gate (sample area) could be mispositioned or of inappropriate size, such that the velocities are overestimates (usually too small) or underestimates (usually too large) of the average velocity. Noncircular cross sections will cause errors in the area estimate, and therefore errors in the flow volume.

Multigate pulsed Doppler systems operate with several parallel channels closely spaced across the lumen of a single large vessel. The outputs from all of the gates can be combined to estimate the velocity profile across the vessel, which represents the variation of flow velocity within the vessel lumen. Velocities mapped with a color scale visually separate the flow information from the gray-scale image, and a real-time color flow Doppler ultrasound image indicates the direction of flow through color coding. However, time is insufficient to complete the computations necessary for determining the Doppler shifts from a large number of gates to get real-time image update rates, particularly for those located at depth.

Doppler Spectral Interpretation

The Doppler signal is typically represented by a spectrum of frequencies resulting from a range of velocities contained within the sampling gate at a specific point in time. Blood flow can exhibit laminar, blunt, or turbulent flow patterns, depending on the vessel wall characteristics, the size and shape of the vessel, and the flow rate. Fast, laminar flow exists in the center of large, smooth wall vessels, while slower blood flow occurs near the vessel walls, due to frictional forces. Turbulent flow occurs at disruptions in the vessel wall caused by plaque buildup and stenosis. A large Doppler gate that is positioned to encompass the entire lumen of the vessel will contain a large range of blood velocities, while a smaller gate positioned in the center of the vessel will have a smaller, faster range of velocities. A Doppler gate positioned near a stenosis in the turbulent flow pattern will measure the largest range of velocities.

With the pulsatile nature of blood flow, the spectral characteristics vary with time. Interpretation of the frequency shifts and direction of blood flow is accomplished with the fast Fourier transform, which mathematically analyzes the detected signals and generates amplitude versus frequency distribution profile known as the Doppler spectrum. In a clinical instrument, the Doppler spectrum is continuously updated in a real-time spectral Doppler display (Fig. 16-50). This

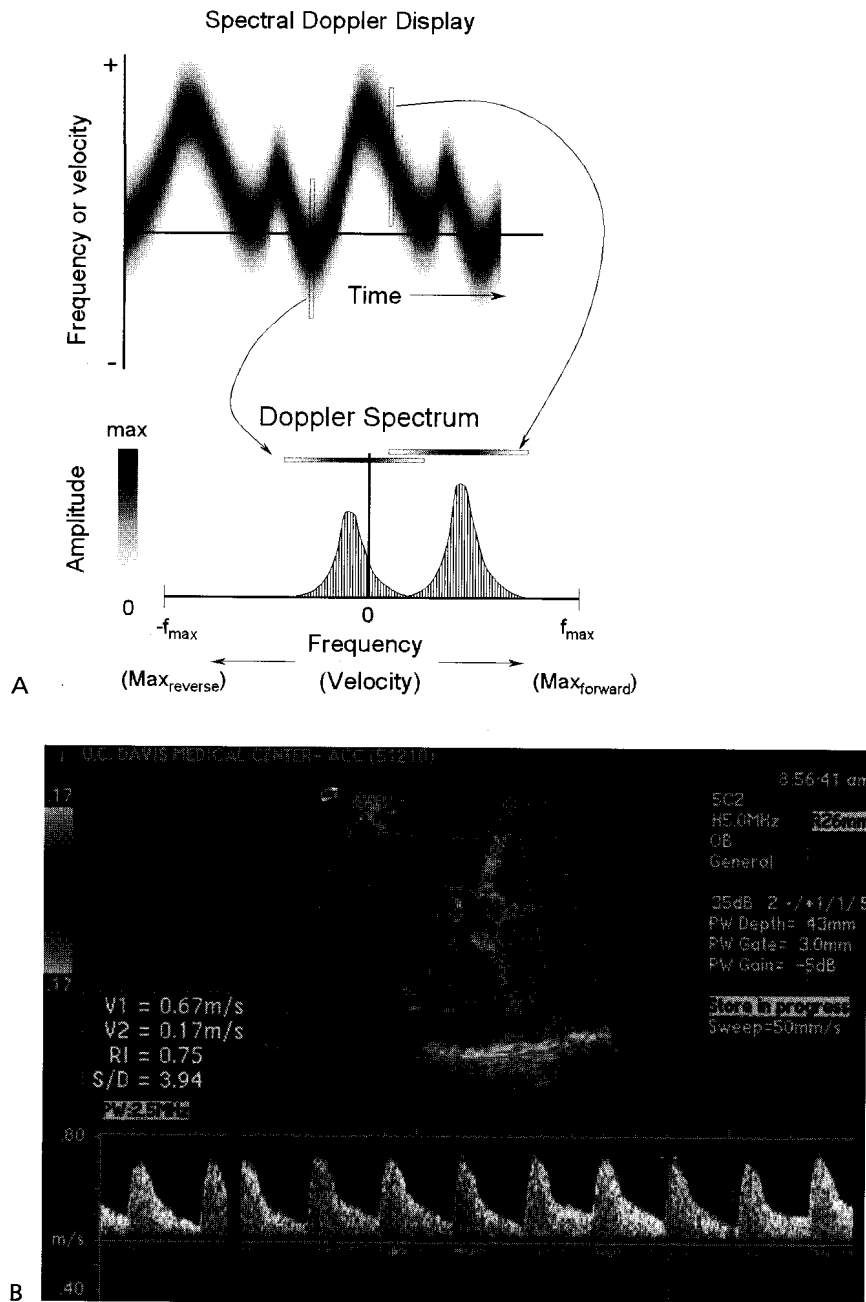


FIGURE 16-50. A: Top graph: The spectral Doppler display is a plot of the Doppler shift frequency spectrum displayed vertically, versus time, displayed horizontally. The amplitude of the shift frequency is encoded as gray-scale variations. **Bottom graph:** Two Doppler spectra are shown from the spectral display at two discrete points in time, with amplitude (gray-scale variations) plotted versus frequency (velocity). A broad spectrum (bandwidth) represents turbulent flow, while a narrow spectrum represents laminar flow within the Doppler gate. **B:** Color Doppler image showing the active color area and the corresponding spectral Doppler display. The resistive index is calculated from the spectral display, and indicated on the image as RI (also see Fig. 16-51).

information is displayed on the video monitor, typically below the 2D B-mode image, as a moving trace, with the blood velocity (proportional to Doppler frequency) as the vertical axis (from $-V_{\max}$ to $+V_{\max}$) and time as the horizontal axis. The intensity of the Doppler signal at a particular frequency and moment in time is displayed as the brightness at that point on the display. Velocities in one direction are displayed as positive values along the vertical axis, and velocities in the other direction are displayed as negative values. As new data arrive, the information is updated and scrolled from left to right. Pulsatile blood takes on the appearance of a choppy sinusoidal wave through the periodic cycle of the heartbeat.

Interpretation of the spectral display provides the ability to determine the presence of flow, the direction of flow, and characteristics of the flow. It is more difficult to determine a lack of flow, since it is also necessary to ensure that the lack of signal is not due to other acoustic or electric system parameters or problems. The direction of flow (positive or negative Doppler shift) is best determined with a small Doppler angle (about 30 degrees). Normal flow is typically characterized by a specific spectral Doppler display waveform, which is a consequence of the hemodynamic features of particular vessels. Disturbed and turbulent flow produce Doppler spectra that are correlated with disease processes. In these latter situations, the spectral curve is "filled in" with a wide distribution of frequencies representing a wide range of velocities, as might occur with a vascular stenosis. Vascular impedance and pulsatile velocity changes concurrent with the circulation can be tracked by Doppler spectrum techniques. Pertinent quantitative measures such as the pulsatility index [$PI = (\max - \min)/\text{average}$] and the resistive index [$RI = (\max - \min)/\max$] are dependent on the characteristics of the Doppler spectral display (Fig. 16-51 describes max, min, and average).

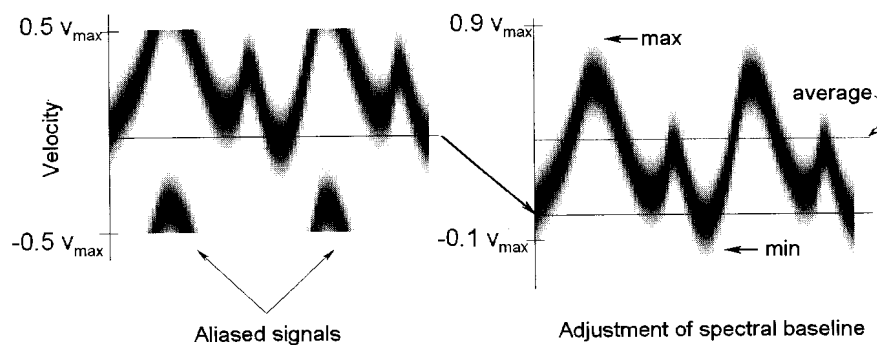


FIGURE 16-51. Left: Aliasing of the spectral Doppler display is characterized by "wrap-around" of the highest velocities to the opposite direction when the sampling (PRF) is inadequate. **Right:** Without changing the overall velocity range, the spectral baseline is shifted to incorporate higher forward velocity and less reverse velocity to avoid aliasing. The maximum, average, and minimum spectral Doppler display values allow quantitative determination of clinically relevant information such as pulsatility index and resistive index.

Aliasing

Aliasing, as described earlier, is an error caused by an insufficient sampling rate (PRF) relative to the high-frequency Doppler signals generated by fast moving blood. A minimum of two samples per cycle of Doppler shift frequency is required to unambiguously determine the corresponding velocity. In a Doppler spectrum, the aliased signals wrap around to negative amplitude, masquerading as reversed flow (Fig. 16-51, left). The most straightforward method to reduce or eliminate the aliasing error is for the user to adjust the velocity scale to a wider range, as most instruments have the PRF of the Doppler unit linked to the scale setting (a wide range delivers a high PRF). If the scale is already at the maximum PRF, the spectral baseline, which represents 0 velocity (0 Doppler shift), can be readjusted to allocate a greater sampling (frequency range) for reflectors moving toward the transducer (Fig. 16-51, right). However, the minimum to the maximum Doppler shift still cannot exceed $\pm \text{PRF}/2$. In such a case, the baseline might be adjusted to $-0.1 V_{\text{max}}$ to $+0.9 V_{\text{max}}$, to allow most of the frequency sampling to be assigned to the positive velocities (positive frequency shifts).

Color Flow Imaging

Color flow imaging provides a 2D visual display of moving blood in the vasculature, superimposed upon the conventional gray-scale image. Velocities and directions are determined for multiple positions within a subarea of the image, and then color encoded (e.g., shades of red for blood moving toward the transducer, and shades of blue for blood moving away from the transducer). Two-dimensional color flow systems do not use the full Doppler shift information because of a lack of time and/or a lack of parallel channels necessary for real-time imaging. Instead, phase-shift autocorrelation or time domain correlation techniques are used.

Phase-shift autocorrelation is a technique to measure the similarity of one scan line measurement to another when the maximum correlation (overlap) occurs. The autocorrelation processor compares the entire echo pulse of one A-line with that of a previous echo pulse separated by a time equal to the pulse repetition period. This “self-scanning” algorithm detects changes in phase between two A-lines of data due to any Doppler shift over the time Δt (Fig. 16-52). The output correlation varies proportionately with the phase change, which in turn varies proportionately with the velocity at the point along the echo pulse trace. In addition, the direction of the moving object (toward or away from the transducer) is preserved through phase detection of the echo amplitudes. Generally, four to eight traces are used to determine the presence of motion along one A-line of the scanned region. Therefore, the beam must remain stationary for short periods of time before insonating another area in the imaging volume. Additionally, because a gray-scale B-mode image must be acquired at the same time, the flow information must be interleaved with the image information. The motion data is mapped with a color scale and superimposed on the gray-scale image. Color flow active area determines the processing time necessary to evaluate the color flow data. A smaller FOV delivers a faster frame rate, but sacrifices the area evaluated for flow. One important consideration is keeping the beam at an angle to the vessel axis at other than 90 degrees. This can be achieved by electronic steering of the color flow ultrasound beams, or by angling the array transducer relative to the vessel axis.

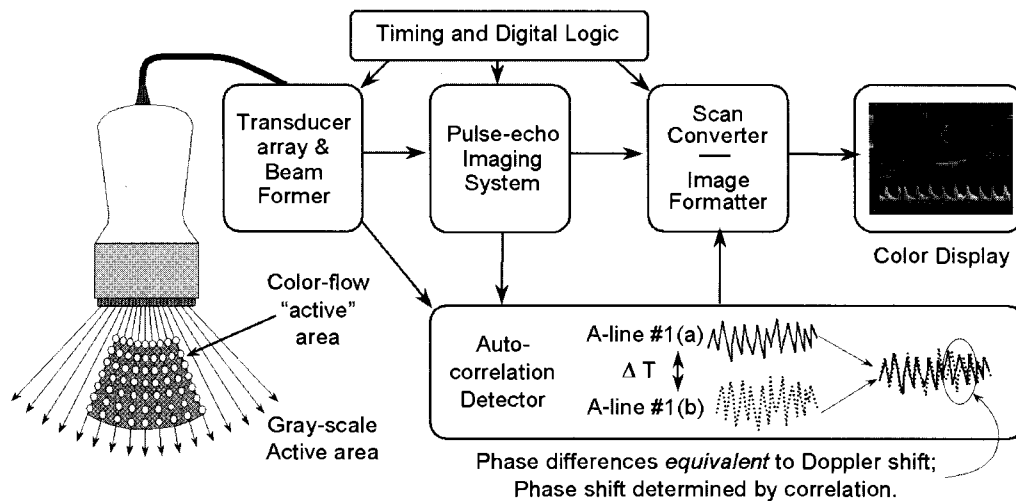


FIGURE 16-52. The color flow imaging system produces dynamic gray-scale B-mode images with color-encoded velocity maps in a user-defined "active area" of multiple "gates." Color assignments depict the direction of the blood (e.g., red toward and blue away from the transducer). Frame rates depend on the number of A-line samples along one position in the image, and the number of lines that compose the active color display area. **Inset:** Autocorrelation detection is a technique to rapidly determine phase changes (equivalent to Doppler shift) in areas of motion deduced from two (or more) consecutive A-lines of data along the same direction.

Time domain correlation is an alternate method for color flow imaging (Fig. 16-53). It is based on the measurement of reflector motion over a time Δt between consecutive pulse echo acquisitions. Correlation mathematically determines the degree of similarity between the two quantities. From echo train one, a series of templates are formed, which are mathematically manipulated over echo train two to determine the time shifts that result in the best correlation. Stationary reflectors need no time shift for high correlation; moving reflectors require a time Δt (either positive or negative) to produce the maximal correlation. The displacement of the reflector (Δx) is determined by the range equation as $\Delta x = (c \Delta t)/2$, where c is the

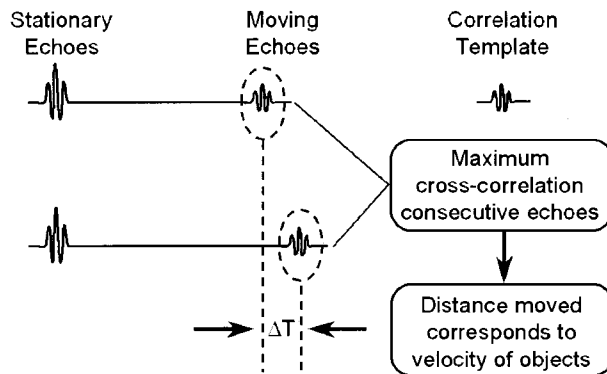


FIGURE 16-53. Time domain correlation uses a short spatial pulse length and an echo "template" to determine positional change of moving reflectors from subsequent echoes. The template scans the echo train to find maximum correlation in each A-line; the displacement between the maximum correlations of each A-line divided by the pulse repetition period is the measured velocity.

speed of sound. Measured velocity (V_m) is calculated as the displacement divided by the time between pulses (the PRP): $V_m = \Delta x / \text{PRP}$. Finally, correction for the angle (θ) between the beam axis and the direction of motion (like the standard Doppler correction) is $V = V_m / \cos(\theta)$. The velocity determines assignment of color, and the images appear essentially the same as Doppler processed images. Multiple pulse echo sequences are typically acquired to provide a good estimate of reflector displacements, but frame update rates are reduced. With time domain correlation methods, short transmit pulses can be used, unlike the longer transmit pulses required for Doppler acquisitions, to achieve narrow bandwidth pulses. This permits better axial resolution. Also, time domain correlation is less prone to aliasing effects compared to Doppler methods because greater time shifts can be tolerated in the returning echo signals from one pulse to the next, which means that higher velocities can be measured.

There are several limitations with color flow imaging. Noise and clutter of slowly moving, solid structures can overwhelm smaller echoes returning from moving blood cells in color flow image. The spatial resolution of the color display is much poorer than the gray-scale image, and variations in velocity are not well resolved in a large vessel. Velocity calculation accuracy by the autocorrelation technique can be limited. Since the color flow map does not fully describe the Doppler frequency spectrum, many color 2D units also provide a duplex scanning capability to provide a spectral analysis of specific questionable areas indicated by the color flow examination. Aliasing artifacts due to insufficient sampling of the phase shifts are also a problem that affects the color flow image, causing apparent reversed flow in areas of high velocity (e.g., stenosis).

Power Doppler

Doppler analysis places a constraint on the sensitivity to motion, because the signals generated by motion must be extracted to determine velocity and direction from the Doppler and phase shifts in the returning echoes within each gated region. In color flow imaging, the frequency shift encodes the pixel value and assigns a color, which is further divided into positive and negative directions. Power Doppler is a signal processing method that relies on the total strength of the Doppler signal (amplitude) and ignores directional (phase) information. The power (also known as energy) mode of signal acquisition is dependent on the amplitude of all Doppler signals, regardless of the frequency shift. This dramatically improves the sensitivity to motion (e.g., slow blood flow) at the expense of directional and quantitative flow information. Compared to conventional color flow imaging, power Doppler produces images that have more sensitivity to motion and are not affected by the Doppler angle (largely nondirectional), and aliasing is not a problem as only the strength of the frequency shifted signals are analyzed (and not the phase). Greater sensitivity allows detection and interpretation of very subtle and slow blood flow. On the other hand, frame rates tend to be slower for the power Doppler imaging mode, and a significant amount of “flash artifacts” occur, which are related to color signals arising from moving tissues, patient motion, or transducer motion. The term *power Doppler* is sometimes mistakenly understood as implying the use of increased transmit power to the patient, but in fact the power levels are typically the same as those in a standard color flow procedure. The difference is in the processing of the returning signals, where sensitivity is achieved at the expense of direction

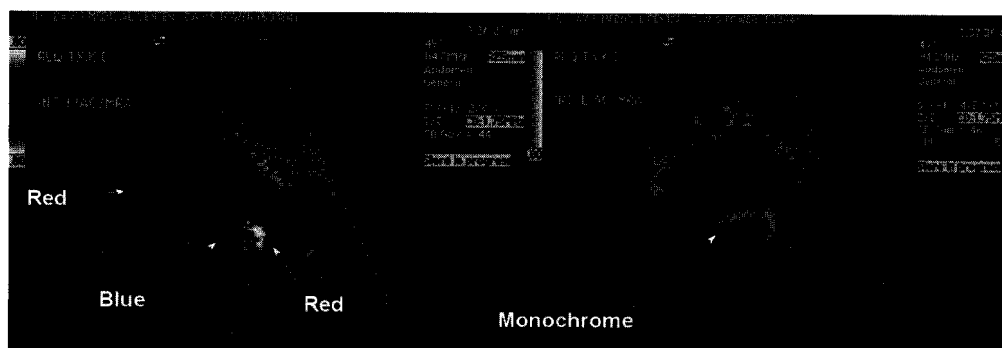


FIGURE 16-54. A comparison of color Doppler (**left**) and power Doppler (**right**) studies show the enhanced sensitivity of the power Doppler acquisition, particularly in areas perpendicular to the beam direction, where the signal is lost in the color Doppler image. Flow directionality, however, is lost in the power Doppler image.

and quantitation. Images acquired with color flow and power Doppler are illustrated in Fig. 16-54.

16.10 SYSTEM PERFORMANCE AND QUALITY ASSURANCE

The system performance of a diagnostic ultrasound unit is described by several parameters: sensitivity and dynamic range, spatial resolution, contrast sensitivity, range/distance accuracy, dead zone thickness, and TGC operation. For Doppler studies, pulse repetition frequency, transducer angle estimates, and range gate stability are key issues. To ensure the performance, accuracy, and safety of ultrasound equipment, periodic quality control (QC) measurements are recommended. The American College of Radiology (ACR) has implemented an accreditation program that specifies recommended periodic quality control procedures for ultrasound equipment. The periodic QC testing frequency of ultrasound components should be adjusted to the probability of finding instabilities or maladjustment. This can be assessed by initially performing tests frequently, reviewing logbooks over an extended period, and, with documented stability, reducing the testing rate.

Ultrasound Quality Assurance

Equipment quality assurance is essentially performed every day during routine scanning by the sonographer, who should and can recognize major problems with the images and the equipment. Ensuring ultrasound image quality, however, requires implementation of a quality control program with periodic measurement of system performance to identify problems before serious malfunctions occur. Tissue-mimicking phantoms are required, having acoustic targets of various sizes and echogenic features embedded in a medium with uniform attenuation and the speed of sound characteristic of soft tissues. Various multipurpose phantoms are available to evaluate the clinical capabilities of the ultrasound system.

A generic phantom composed of three modules is illustrated in Fig. 16-55. The phantom gel filler has tissue-like attenuation of 0.5 to 0.7 (dB/cm)/MHz (higher attenuation provides a more challenging test) and low-contrast targets within a

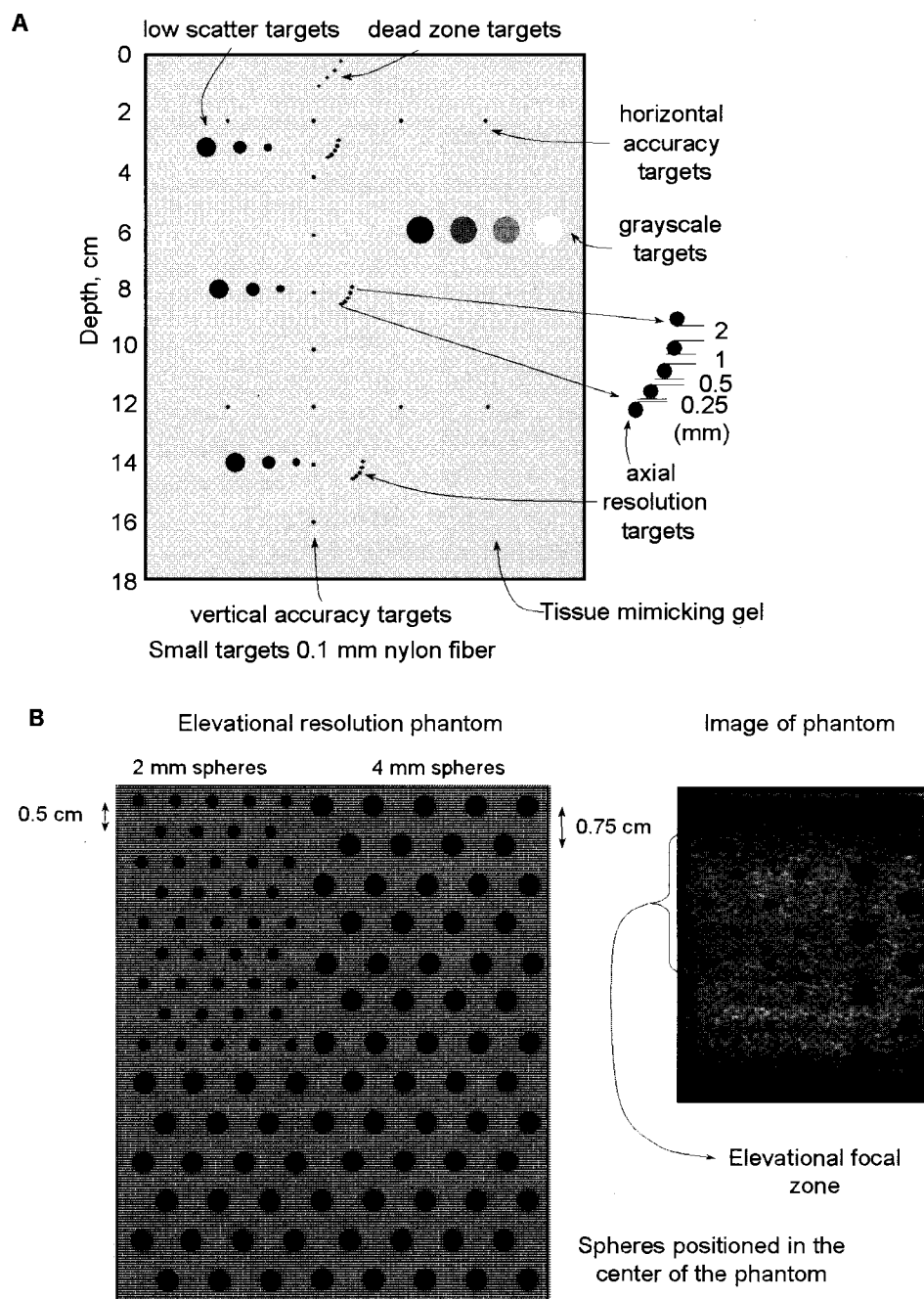


FIGURE 16-55. General-purpose ultrasound quality assurance phantoms are typically composed of several scanning modules. **A:** System resolution targets (axial and lateral), dead zone depth, vertical and horizontal distance accuracy targets, contrast resolution (gray-scale targets), and low-scatter targets positioned at several depths (to determine penetration depths) are placed in a tissue mimicking (acoustic scattering) gel. **B:** Elevational resolution is determined with spheres equally distributed along a plane in tissue mimicking gel. An image of the phantom shows the effects of partial volume averaging and variations in elevational resolution with depth. (Image reprinted by permission of Gammex, Inc., Madison, WI.)

(continued on next page)

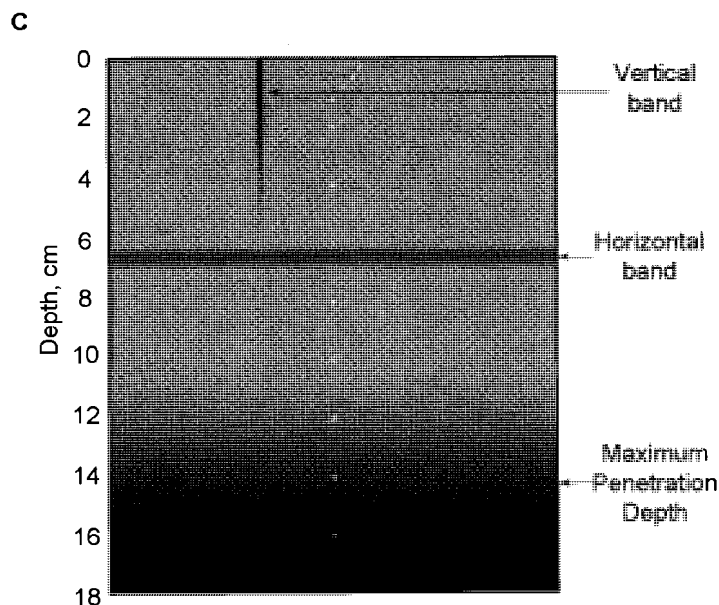


FIGURE 16-55 (continued). C: The system uniformity module elucidates possible problems with image uniformity. Shown here is a simulation of a transducer malfunction (vertical dropout), and horizontal transition mismatch. Penetration depth for uniform response can also be determined.

matrix of small scatterers to mimic tissue background. Small, high-contrast reflectors are positioned at known depths for measuring the axial and lateral spatial resolution, for assessing the accuracy of horizontal and vertical distance measurements, and for measuring the depth of the dead zone (the nonimaged area immediately adjacent to the transducer). Another module contains low-contrast, small diameter spheres (or cylinders) of 2- and 4-mm diameter uniformly spaced with depth, to measure elevational resolution (slice thickness) variation with depth. The third module is composed of a uniformly distributed scattering material for testing image uniformity and penetration depth.

Spatial resolution, contrast resolution, and distance accuracy are evaluated with one module (Fig. 16-55A). Axial resolution is evaluated by the ability to resolve high-contrast targets separated by 2, 1, 0.5, and 0.25 mm at three different depths. In an optimally functioning system, the axial resolution should be consistent with depth and improve with higher operational frequency. Lateral resolution is evaluated by measuring the lateral spread of the high-contrast targets as a function of depth and transmit focus. Contrast resolution is evaluated with “gray-scale” objects of lower and higher attenuation than the tissue mimicking gel; more sophisticated phantoms have contrast resolution targets of varying contrast and size. Contrast resolution should improve with increased transmit power. Dead zone depth is determined with the first high-contrast target (positioned at several depths from 0 to ~1 cm) visible in the image. Horizontal and vertical distance measurement accuracy is determined with the small high-contrast targets. Vertical targets (along the axial beam direction) should have higher precision and accuracy than the corresponding

horizontal targets (lateral resolution), and all measurements should be within 5% of the known distance.

Elevational resolution and partial volume effects are evaluated with the “sphere” module (Fig. 16-55B). The ultrasound image of the spherical targets illustrates the effects of slice thickness variation with depth. With the introduction of 1.5-D transducer arrays as well as 3-D imaging capabilities, multiple transmit focal zones to reduce slice thickness at various depths are becoming important, as is the need to verify elevational resolution performance.

Uniformity and penetration depth is measured with the uniformity module (Fig. 16-55C). With a properly adjusted and operating ultrasound system, a uniform response is expected up to the penetration depth capabilities of the transducer array. Higher operational frequencies will display a lower penetration depth. Evidence of vertically directed shadowing for linear arrays or angular directed shadowing for phased arrays is an indication of malfunction of a particular transducer element or its associated circuitry. Horizontal variations indicate inadequate handling of transitions between focal zones when in the multifocal mode. These two problems are simulated in the figure of the uniformity module.

During acceptance testing of a new unit, all transducers should be evaluated, and baseline performance measured against manufacturer specifications. The maximum depth of visualization is determined by identifying the deepest low-contrast scatterers in a uniformity phantom that can be perceived in the image. This depth is dependent on the type of transducer and its operating frequency, and should be measured to verify that the transducer and instrument components are operating at their designed sensitivity levels. In a 0.7 dB/cm-MHz attenuation medium, a depth of 18 cm for abdominal and 8 cm for small-parts transducers is the goal for visualization when a single transmit focal zone is placed as deeply as possible with maximum transmit power level and optimal TGC adjustments. For a multifrequency transducer, the midfrequency setting should be used. With the same settings, the uniformity section of the phantom without targets assesses gray level and image uniformity. Power and gain settings of the machine can have a significant effect on the apparent size of point-like targets (e.g., high receive gain reveals only large-size targets and has poorer resolution). One method to improve test reproducibility is to set the instrument to the threshold detectability of the targets, and to rescan with an increased transmit gain of 20 dB.

Recommended QC procedures for ACR accreditation are listed in Table 16-8. Routine QC testing must occur regularly; a minimum requirement for accreditation is semiannually. The same tests must be performed during each testing period so that changes can be monitored over time and effective corrective action can be taken. Testing results, corrective action, and the effects of corrective action must be documented and maintained on site. Other equipment-related issues involve cleaning air filters, checking for loose or frayed cables, and checking handles, wheels, and wheel locks as part of the QC tests.

The most frequently reported source of performance instability of an ultrasound system is related to the recording of the images on film or the display on maladjusted video monitors. Drift of the ultrasound instrument settings or film camera settings, poor film processing conditions, and/or poor viewing conditions (for instance, portable ultrasound performed in a very bright patient room, causing inappropriately gain-adjusted images) can lead to suboptimal images on both hard-copy (film) and softcopy (monitor). The analog contrast and brightness settings for

TABLE 16-8. RECOMMENDED QC TESTS FOR AMERICAN COLLEGE OF RADIOLOGY (ACR) ACCREDITATION PROGRAM

Test (Gray Scale Imaging Mode) for Each scanner	Minimum Frequency
System sensitivity and/or penetration capability	Semiannually
Image uniformity	Semiannually
Photography and other hard copy	Semiannually
Low-contrast detectability (optional)	Semiannually
Assurance of electrical and mechanical safety	Semiannually
Horizontal and vertical distance accuracy	At acceptance
Transducers (of different scan format)	Ongoing basis

the monitor and multifORMAT camera should be properly established during installation, and should not be routinely adjusted. Properly adjusting the video monitor brightness/contrast knobs can be accomplished with image test patterns (e.g., the SMPTE pattern; see Chapter 4). A maximum window width and middle window level displays the minimum to the maximum gray-scale values, and allows the analog monitor contrast and brightness controls to be properly adjusted. Once adjusted, the analog knobs must be disabled so that image scaling is only applied using window/level tuning from the scan converter memory. The image-recording (film) device is then adjusted to the same range of gray-scale values that appear on the soft copy device.

Doppler Performance Measurements

Doppler techniques are becoming more common in the day-to-day use of medical ultrasound equipment. Reliance on flow measurements to make diagnoses requires demonstration of accurate data acquisition and processing. Quality control phantoms to assess velocity and flow contain one or more tubes in tissue mimicking materials at various depths. A blood-mimicking fluid is pushed through the tubes with carefully calibrated pumps to provide a known velocity for assessing the accuracy of the Doppler velocity measurement. Several tests can be performed, including maximum penetration depth at which flow waveforms can be detected, alignment of the sample volume with the duplex B-mode image, accuracy of velocity measurements, and volume flow. For color-flow systems, sensitivity and alignment of the color flow image with the B-scan gray-scale image are assessed.

16.11 ACOUSTIC POWER AND BIOEFFECTS

Power is the rate of energy production, absorption, or flow. The SI unit of power is the watt (W), defined as 1 joule of energy per second. Acoustic intensity is the rate at which sound energy flows through a unit area and is usually expressed in units of watts per square centimeter (W/cm^2) or milliwatts per square centimeter (mW/cm^2).

Ultrasound acoustic power levels are strongly dependent on the operational characteristics of the system, including the transmit power, pulse repetition frequency, transducer frequency, and operation mode. Biologic effects (bioeffects) are predominately related to the heating of tissues caused by high-intensity levels of

ultrasound used to enhance image quality and functionality. For diagnostic imaging, the intensity levels are kept below the threshold for established bioeffects.

Acoustic Power and Intensity of Pulsed Ultrasound

Measurement Methods

Measurement of ultrasound pressure amplitude within a beam is performed with a hydrophone, a device containing a small (e.g., 0.5-mm-diameter) piezoelectric element coupled to external conductors and mounted in a protective housing. When placed in an ultrasound beam, the hydrophone produces a voltage that is proportional to the variations in pressure amplitude at that point in the beam as a function of time, permitting determination of peak compression and rarefaction amplitude as well as pulse duration and pulse repetition period (Fig. 16-56A). Calibrated hydrophones provide absolute measures of pressure, from which the acoustic intensity can be calculated if the acoustic impedance of the medium is accurately known.

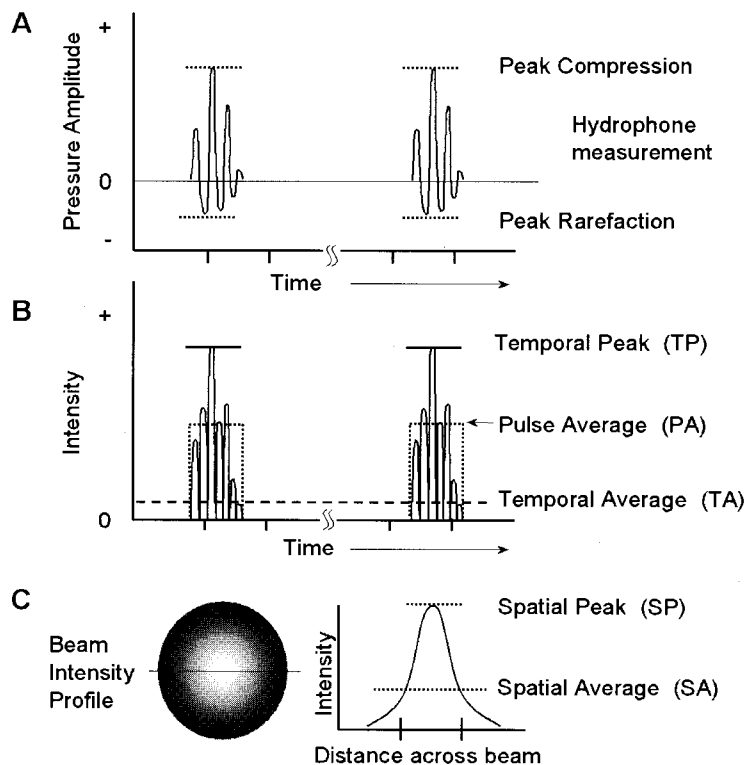


FIGURE 16-56. A: Pressure amplitude variations are measured with a hydrophone, and include peak compression and peak rarefaction variations with time. **B:** Temporal intensity variations of pulsed ultrasound vary widely, from the temporal peak and temporal average values; pulse average intensity represents the average intensity measured over the pulse duration. **C:** Spatial intensity variations of pulsed ultrasound are described by the spatial peak value and the spatial average value, measured over the beam profile.

Intensity Measures of Pulsed Ultrasound

In the pulsed mode of ultrasound operation, the instantaneous intensity varies greatly with time and position. At a particular location in tissue, the instantaneous intensity is quite large while the ultrasound pulse passes through the tissue, but the pulse duration is only about a microsecond or less, and for the remainder of the pulse repetition period the intensity is nearly zero.

The temporal peak, I_{TP} , is the highest instantaneous intensity in the beam; the temporal average, I_{TA} , is the time-averaged intensity over the pulse repetition period; and the pulse average, I_{PA} , is the average intensity of the pulse (Fig. 16-56B). The spatial peak, I_{SP} , is the highest intensity spatially in the beam, and the spatial average, I_{SA} , is the average intensity over the beam area, usually taken to be the area of the transducer (Fig. 16-56C).

The acoustic power contained in the ultrasound beam (watts), averaged over at least one pulse repetition period, and divided by the beam area (usually the area of the transducer face) is the spatial average–temporal average intensity I_{SATA} . Other meaningful measures for pulsed ultrasound intensity are determined from I_{SATA} , including the following:

1. The spatial average–pulse average intensity, $I_{SAPA} = I_{SATA}/\text{duty cycle}$, where $I_{PA} = I_{TA}/\text{duty cycle}$.
2. The spatial peak–temporal average intensity, $I_{SPTA} = I_{SATA} (I_{SP}/I_{SA})$, which is a good indicator of thermal ultrasound effects.
3. The spatial peak–pulse average intensity, $I_{SPPA} = I_{SATA} (I_{SP}/I_{SA})/\text{duty cycle}$, an indicator of potential mechanical bioeffects and cavitation.

For acoustic ultrasound intensity levels, $I_{SPPA} > I_{SPTA} > I_{SAPA} > I_{SATA}$. Typical acoustical power outputs are listed in Table 16-9. The two most relevant measures are the I_{SPPA} and the I_{SPTA} . Both measurements are required by the U.S. Food and Drug Administration (FDA) for certification of instrumentation. Values of I_{SPTA} for diagnostic imaging are usually below 100 mW/cm^2 for imaging, but for certain Doppler applications I_{SPTA} can exceed $1,000 \text{ mW/cm}^2$. I_{SPPA} can be substantially greater than I_{SPTA} , as shown in Table 16-8. For real-time scanners, the combined intensity descriptors must be modified to consider dwell time (the time the ultrasound beam is directed at a particular region), and the acquisition geometry and spatial sampling. These variations help explain the measured differences between the I_{SPTA} and I_{SPPA} values indicated, which are much less than the duty cycle values that are predicted by the equations above.

TABLE 16-9. TYPICAL INTENSITY MEASURES FOR ULTRASOUND DATA COLLECTION MODES

Mode	Pressure Amplitude (MPa)	I_{SPTA} (mW/cm ²)	I_{SPPA} (W/cm ²)	Power (mW)
B-scan	1.68	19	174	18
M-mode	1.68	73	174	4
Pulsed doppler	2.48	1,140	288	31
Color flow	2.59	234	325	81

Adapted from compilation of data presented by the American Institute of Ultrasound in Medicine.
Note the difference in units for I_{SPTA} (mW/cm²) versus I_{SPPA} (W/cm²).

Thermal and mechanical indices of ultrasound operation are now the accepted method of determining power levels for real-time instruments that provide the operator with quantitative estimates of power deposition in the patient. These indices are selected for their relevance to risks from biologic effects, and are displayed on the monitor during real-time scanning. The sonographer can use these indices to minimize power deposition to the patient (and fetus) consistent with obtaining useful clinical images in the spirit of the ALARA (as low as reasonably achievable) concept.

Thermal Index

The thermal index, TI, is the ratio of the acoustic power produced by the transducer to the power required to raise tissue in the beam area by 1°C. This is estimated by the ultrasound system using algorithms that take into account the ultrasonic frequency, beam area, and the acoustic output power of the transducer. Assumptions are made for attenuation and thermal properties of the tissues with long, steady exposure times. An indicated TI value of 2 signifies a possible 2°C increase in the temperature of the tissues when the transducer is stationary. TI values are associated with the I_{SPTA} measure of intensity.

On some scanners, other thermal indices that might be encountered are TIS (S for soft tissue), TIB (B for bone), and TIC (C for cranial bone). These quantities are useful because of the increased heat buildup that can occur at a bone–soft tissue interface when present in the beam, particularly for obstetrics scanning of late-term pregnancies, and with the use of Doppler ultrasound (where power levels can be substantially higher).

Mechanical Index

Cavitation is a consequence of the negative pressures (rarefaction of the mechanical wave) that induce bubble formation from the extraction of dissolved gases in the medium. The mechanical index (MI) is a value that estimates the likelihood of cavitation by the ultrasound beam. The MI is directly proportional to the peak rarefactional (negative) pressure, and inversely proportional to the square root of the ultrasound frequency (in MHz). An attenuation of 0.3 (dB/cm)/MHz is assumed for the algorithm that estimates the MI. As the ultrasound output power (transmit pulse amplitude) is increased, the MI increases linearly, while an increase in the transducer frequency (say from 2 to 8 MHz) decreases the MI by the square root of 4, or by a factor of two. MI values are associated with the I_{SPPA} measure of intensity.

Biologic Mechanisms and Effects

Diagnostic ultrasound has established a remarkable safety record. Significant deleterious bioeffects on either patients or operators of diagnostic ultrasound imaging procedures have not been reported in the literature. Despite the lack of evidence that any harm can be caused by diagnostic intensities of ultrasound, it is prudent and indeed an obligation of the physician to consider issues of benefit versus risk when performing an ultrasound exam, and to take all precautions to ensure maximal benefit with minimal risk. The American Institute of Ultrasound in Medicine

recommends adherence to the ALARA principles. FDA requirements for new ultrasound equipment include the display of acoustic output indices (MI and TI) to give the user feedback regarding the power deposition to the patient.

At high intensities, ultrasound can cause biologic effects by thermal and mechanical mechanisms. Biologic tissues absorb ultrasound energy, which is converted into heat; thus, heat will be generated at all parts of the ultrasonic field in the tissue. Thermal effects are dependent not only on the rate of heat deposition in a particular volume of the body, but also on how fast the heat is removed by blood flow and other means of heat dissipation. The best indicator of heat deposition is the I_{SPTA} measure of intensity and the calculated TI value. Heat deposition is determined by the average ultrasound intensity in the focal zone and the absorption coefficient of the tissue. Absorption increases with the frequency of the ultrasound and varies with tissue type. Bone has a much higher attenuation (absorption) coefficient than soft tissue, which can cause significant heat deposition at a tissue-bone interface. In diagnostic ultrasound applications, the tissue temperature rise (e.g., 1° to 2°) is typically well below that which would be considered potentially damaging, although some Doppler instruments can approach these levels with high pulse repetition frequencies and longer pulse duration.

Nonthermal mechanisms include mechanical movement of the particles of the medium due to radiation pressure (which can apply force or torque on tissue structures) and acoustic streaming, which can give rise to a steady circulatory flow. With higher energy deposition over a short period, cavitation can occur. Cavitation can be broadly defined as sonically generated activity of highly compressible bodies composed of gas and/or vapor. Cavitation can be subtle or readily observable, and is typically unpredictable and sometimes violent. Stable cavitation generally refers to the pulsation (expansion and contraction) of persistent bubbles in the tissue that occur at low and intermediate ultrasound intensities (as used clinically). Chiefly related to the peak rarefactional pressure, the MI is an estimate for producing cavitation. At higher ultrasound intensity levels, transient cavitation can occur, whereby the bubbles respond nonlinearly to the driving force, causing a collapse approaching the speed of sound. At this point, the bubbles might dissolve, disintegrate, or rebound. In the minimum volume state, conditions exist that can dissociate the water vapor into free radicals such as $H\cdot$ and $OH\cdot$, which can cause chemical damage to biologically important molecules such as DNA. Short pulses such as those used in imaging are good candidates for transient cavitation; however, the intensities used in diagnostic imaging are far below the transient cavitation threshold (e.g., 1 kW/cm^2 peak pulse power is necessary for transient cavitation to be evoked).

Although biologic effects have been demonstrated at much higher ultrasound power levels and longer durations, the levels and durations for typical imaging and Doppler studies are below the threshold for known adverse effects. At higher output power, outcomes include macroscopic damage (e.g., rupturing of blood vessels, breaking up cells—indeed, the whole point of shock wave lithotripsy is the breakup of kidney stones) and microscopic damage (e.g., breaking of chromosomes, changes in cell mitotic index). No bioeffects have been shown below an I_{SPTA} of 100 mW/cm^2 (Fig. 16-57). Even though ultrasound is considered safe when used properly, prudence dictates that diagnostic ultrasound exposure be limited to only those patients for whom a definite benefit will be obtained.

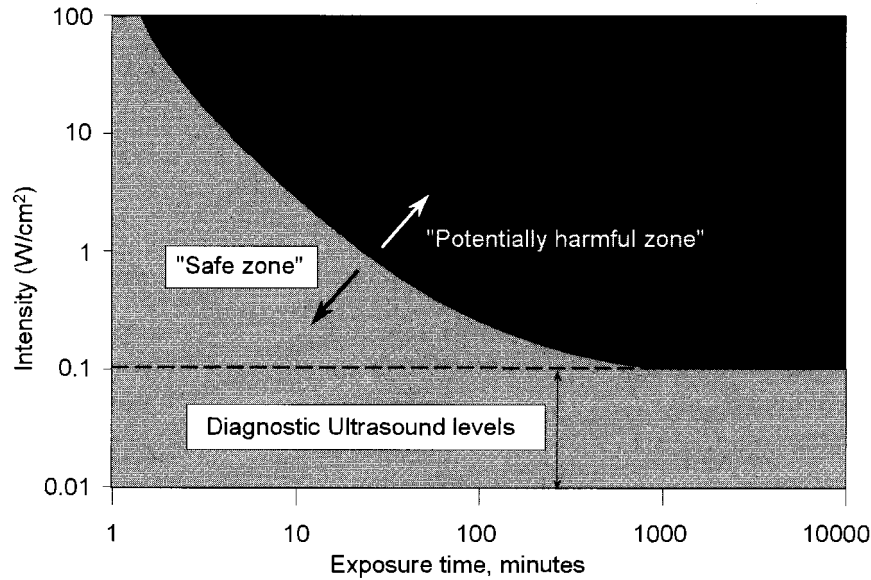


FIGURE 16-57. The potential bioeffects from ultrasound. Safe and potentially harmful regions are delineated according to ultrasound intensity levels and exposure time. The dashed line shows the upper limit of intensities typically encountered in diagnostic imaging applications.

SUGGESTED READING

- American Institute of Ultrasound in Medicine. Official statements and reports. Available at the AIUM Web site: <http://www.aium.org>.
- Hedrick WR, Hykes DL, Starchman DE. *Ultrasound physics and instrumentation*, 3rd ed. St. Louis: Mosby-Year Book, 1995.
- Kremkau FW. *Diagnostic ultrasound: principles and instruments*, 5th ed. Philadelphia: WB Saunders, 1998.
- Zagzebski JA. *Essentials of ultrasound physics*. St. Louis: Mosby-Year Book, 1996.

COMPUTER NETWORKS, PACS, AND TELERADIOLOGY

INTRODUCTION

Advances in digital imaging technology and the rapidly falling costs and increasing capabilities of computer networks, workstations, storage media, and display devices have spurred a considerable interest in picture archiving and communications systems (PACS) and teleradiology, with the goal of improving the utilization and efficiency of radiologic imaging. For example, a PACS can replicate images at different display workstations simultaneously for the radiologist and referring physician and can be a repository for several years worth of images. The loss of films, a major problem in teaching hospitals, can be essentially eliminated. The transfer of images via teleradiology can bring subspecialty expertise to rural areas and provide around-the-clock trauma radiology services. Knowledge of the basic aspects of these systems is important for their optimal use and is the topic of this chapter. Computer networks, which comprise the information pathways linking the other components of PACS and teleradiology systems, are discussed in depth in the first part of this chapter. The remainder of the chapter is devoted to the components and functioning of PACS and teleradiology.

17.1 COMPUTER NETWORKS

Basic Principles

Computer networks permit the transfer of information between computers; allow computers to share devices, such as printers and laser multifunction cameras; and support services such as electronic transmission of messages (e-mail), transfer of computer files, and use of distant computers. Networks, based on the distances they span, may be described as local area networks or wide area networks. A *local area network* (LAN) connects computers within a department, a building such as a medical center, and perhaps neighboring buildings, whereas a *wide area network* (WAN) connects computers at large distances from each other. LANs and WANs evolved separately, and a computer may be connected to a WAN without being part of a LAN. However, most WANs today consist of multiple LANs connected by medium- or long-distance communication links. The largest WAN is the Internet.

Networks have both hardware and software components. A physical connection must exist between computers so that they can exchange information. Common physical connections include coaxial cable, copper wiring, and optical fiber cables, but they may also include microwave and satellite links. A coaxial cable is a

cylindrical cable with a central conductor surrounded by an insulator that, in turn, is surrounded by a tubular grounded shield conductor. Coaxial cable and copper wiring carry electrical signals. Optical fiber cables use glass or plastic fibers to carry light signals produced by lasers or light-emitting diodes. Optical fiber cables have the advantage that they are not affected by electrical interference and therefore usually have lower error rates than cables carrying electrical signals. There are also layers of software between the application program (application) with which the user interacts and the hardware of the communications link.

The data transfer rate of a computer network is usually described in megabits per second (10^6 bps = 1 Mbps) or gigabits per second (10^9 bps = 1 Gbps). These units should not be confused with megabytes per second (MBps) and gigabytes per second (GBps), commonly used to specify the data transfer rates of computer components such as magnetic and optical disks. (Recall that a byte consists of eight bits.) A further complication is that kilobytes, megabytes, and gigabytes, when used to describe the capacity of computer memory or a storage device, are defined in powers of two: 1 kB = 2^{10} bytes = 1,024 bytes, 1 MB = 2^{20} bytes = 1,048,576 bytes, and 1 GB = 2^{30} bytes = 1,073,741,824 bytes (see Table 4.3 in Chapter 4). The maximal data transfer rate is often called the *bandwidth*, a term originally used to describe the data transfer rates of analog communications channels. An actual network may not achieve its full nominal bandwidth because of overhead or inefficiencies in its implementation. The term *throughput* is often used to describe the maximal data transfer rate that is actually achieved.

In most networks, multiple computers share communication pathways. Most network protocols facilitate this sharing by dividing the information to be transmitted into *packets*. Some protocols permit packets of variable size, whereas others permit only packets of a fixed size. Packets may be called frames, datagrams, or cells. Each packet contains information identifying its destination. In most protocols, a packet also identifies the sender and contains information used to determine whether the packet's contents were garbled during transfer. The computer at the final destination reassembles the information from the packets and may request retransmission of lost packets or packets with errors.

Large networks often employ switching devices to forward packets between network segments or even between entire networks. Each device on a network, whether a computer or switching device, is called a *node*, and the communications pathways between them are called *links*. Each computer is connected to a network by a network adapter, commonly called a network interface card, installed on the input/output (I/O) bus of the computer (see Chapter 4). Each interface between a node and a network is identified by a unique number called a *network address*. Computers usually have only a single interface, but a switching device connecting two or more networks may have an address on each network.

On many small LANs today, all computers are directly connected—that is, all packets reach every computer. However, on a larger network, every packet reaching every computer would be likely to cause unacceptable network congestion. For this reason, most networks larger than a small LAN, and even some small LANs, employ *packet switching*. The packets are sent over the network. Devices called bridges, switches, or routers (discussed later) store the packets, read the destination addresses, and send them on toward their destinations (a method called “store and forward”). In some very large packet-switched networks, individual packets from one computer to another may follow different paths through the network and can arrive out of order.

Network protocols describe the methods by which communication occurs over a network. Both hardware and software must comply with established communica-

tions protocols. Failure to conform to a common protocol would be analogous to a person speaking only English attempting to converse with a person speaking only Chinese. Several hardware and software protocols are described later in this chapter.

Network protocols can be divided into several layers (Fig. 17-1). The lowest layer, the Physical Layer, describes the physical circuit (the physical wiring or optical fiber cable connecting nodes and the electrical or light signals sent between nodes). Layer 2, the Data Link Layer, describes the LAN packet format and functions such as media access control (i.e., determining when a node may transmit a packet on a LAN) and error-checking of packets received over a LAN. It is usually implemented in hardware. Layers 1 and 2 are responsible for transmission of packets from one node to another over a LAN and enable computers with dissimilar hardware and operating systems to be physically connected. However, they alone do not permit applications to communicate across a network. There must be higher-level network protocols to mediate between application programs and the network interface adapter. These protocols usually are implemented in software and incorporated in a computer's operating system. Many higher-level protocols are available, their complexity depending on the scope and complexity of the networks they are designed to serve.

Computer networks allow applications on different computers to communicate. Network communications begin at the Application Layer (Layer 5 in Fig. 17-1) on a computer. The application passes the information to be transmitted to the next lower layer in the stack. The information is passed from layer to layer, with each layer adding information, such as addresses and error-detection information, until it reaches the Physical Layer. The Physical Layer sends the information to the destination computer, where it is passed up the protocol layer stack to the application layer of the destination computer.

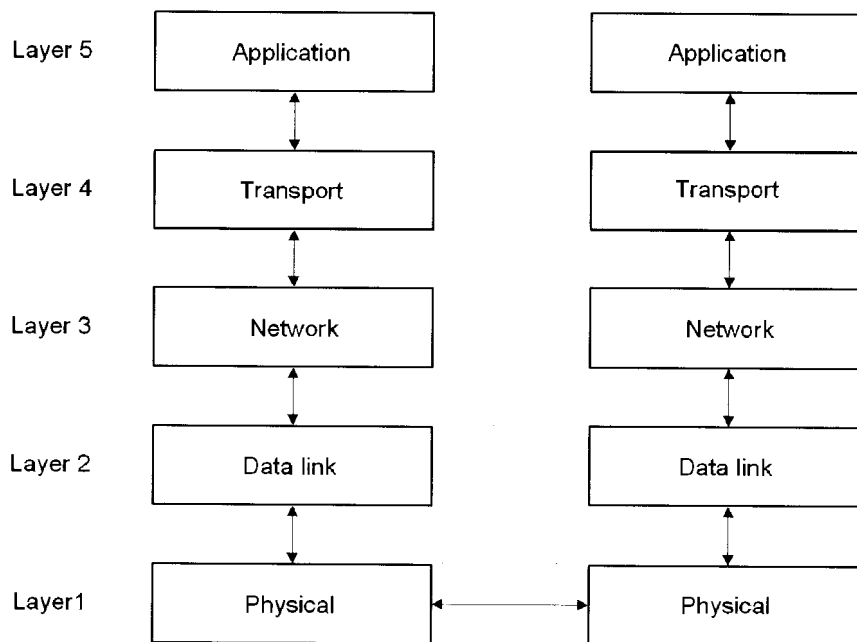


FIGURE 17-1. Transmission Control Protocol/Internet Protocol (TCP/IP) Network Protocol Stack. A more complex network model, the Open Systems Interconnection (OSI) model, is also commonly used to model networks. The OSI model has seven layers. The bottom four layers of the OSI model match the bottom four layers shown in this figure. However, the OSI model divides the Application Layer of the TCP/IP Protocol Stack into three separate layers.

On most networks today, when two nodes communicate over a link, only one node sends information while another receives it. This is called *half-duplex operation*. *Full-duplex* operation occurs when two nodes send and receive simultaneously.

The term *server* is used to refer to a computer on a network that provides a service to other computers on the network. A computer with a large array of magnetic disks that provides data storage for other computers is called a file server. There are also print servers, application servers, database servers, e-mail servers, web servers, and so on.

Local Area Networks

A LAN can connect computers and other devices only over a limited distance without devices such as repeaters to extend the network. On many small LANs, the computers are all directly connected; only one computer can transmit at a time, and usually only a single computer accepts the information. This places a practical limit on the number of computers and other devices that can be placed on a LAN without excessive network congestion. The congestion caused by too many devices on a LAN can be relieved by dividing the LAN into *segments* connected by intelligent devices, such as bridges, switches, or routers, that transmit only information intended for other segments.

Most types of LANs are configured in bus, star, or ring topologies (Fig. 17-2). In a *bus topology*, all of the nodes are connected to a single cable. One node sends a packet, including the network address of the destination node, and the destination node accepts the packet. All other nodes must wait. In a *star topology*, all nodes are connected to a central node, often called a hub. In a *ring topology*, each node is directly connected to only two adjacent nodes, but all of the nodes and their links form a ring.

Any network on which multiple nodes share links must have a *media access control protocol* specifying when nodes can send information. Most bus networks are contention-based, meaning that nodes preparing to transmit contend for the use of the bus. Most ring networks employ a method called *token passing* to control access. A special bit sequence, called a *token*, is sent from each node to the next around the ring, until it reaches a node that is waiting to transmit. A node can transmit only if it holds the token. The node holding the token transmits a packet to the next node on the ring. The packet is received and retransmitted by each node around the ring until it reaches the destination node. Once a node has finished transmitting, it sends the token to the next node, offering that node the opportunity to transmit.

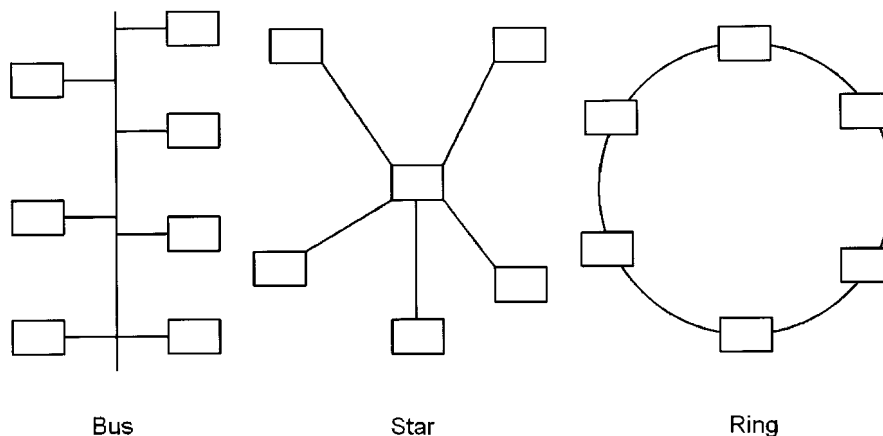


FIGURE 17-2. Network topologies—bus, star, and ring.

Ethernet

The most common LAN medium is Ethernet. Before transmission over Ethernet, information to be transmitted is divided into packets, each with a header specifying the addresses of the transmitting and destination nodes. Ethernet is contention-based. A node ready to transmit a packet first “listens” to determine whether another node is transmitting. If none is, it attempts to transmit. If two nodes inadvertently attempt to transmit at the same moment, a collision occurs. (In Ethernet, collisions are normal expected events.) When a collision occurs, each node ceases transmission, waits for a randomly determined but traffic-dependent time interval, and again attempts to transmit. This media access control protocol is called *carrier-sense multiple-access with collision detection* (CSMA/CD).

There are many varieties of Ethernet (Table 17-1). 10Base-5 and 10Base-2 Ethernet use the bus topology, with all nodes connected to a coaxial cable. In such clas-

TABLE 17-1. COMMON LOCAL AREA NETWORKS^a

Type	Data transfer rate (Mbps)	Physical topology	Logical topology	Medium	Maximal range
Ethernet (10Base-5)	10	Bus	Bus	Coaxial cable	500 m
Ethernet (10Base-2)	10	Bus	Bus	Thin coaxial cable	185 m
Ethernet (10Base-T)	10	Star	Bus	UTP (common telephone wiring)	100 m from node to hub
Fast Ethernet (100Base-TX)	100	Star	Bus	Category 5 UTP	100 m from node to hub
Switched Ethernet (10Base-T)	10 between each pair of nodes	Star	Star	UTP (common telephone wiring)	100 m from node to switch
Switched Fast Ethernet (100Base-TX)	100 between each pair of nodes	Star	Star	Category 5 UTP	100 m from node to switch
Gigabit Ethernet (1000Base-CX and 1000Base-T)	1,000	Star	Star	Copper wiring	At least 25 m and perhaps 100 m
Gigabit Ethernet (1000Base-SX and 1000Base-LX)	1,000	Star	Star	Optical fiber cable	550 m (multimode cable) and 3 km (single mode)
Fiber Distributed Data Interface (FDDI)	100	Ring	Ring	Optical fiber cable	2 km (multimode) and 40 km (single mode) between each pair of nodes
Asynchronous Transfer Mode (ATM)	1.5, 25, 100, 155, 622, and 2,488	—	—	Various	—

Note: UTP, unshielded twist pair.

^a“Physical topology” describes the actual topology of the medium, whereas “logical topology” describes how it functions. For example, unswitched 10Base-T Ethernet is wired in a star topology, with each computer connected to a central hub by a length of cable. However, it behaves like a bus topology, because the hub “broadcasts” each packet to all nodes connected to the hub.

sis (nonswitched) forms of Ethernet, the transmitting node “broadcasts” each packet. All other nodes receive the packet and read its destination address, but only the destination node stores the information in memory. 10Base-T Ethernet is physically configured in a star topology, with each node connected to a central hub by unshielded twisted pair (UTP) telephone wiring. However, 10Base-T Ethernet behaves logically as a bus topology, because the hub broadcasts the packets to all nodes. 10Base-5, 10Base-2, and 10Base-T Ethernet support data transfer rates up to 10 Mbps. Fast Ethernet (100Base-TX) is like 10Base-T, except that it permits data transfer rates up to 100 Mbps over high-grade Category 5 UTP wiring.

A 10Base-T or Fast Ethernet network can be converted to a switched Ethernet network by replacing the hub with an Ethernet switch. The switch, in contrast to a hub, does not broadcast the packets to all nodes. Instead, it reads the address on each packet and forwards the packet only to the destination node. In this way, the switch permits several pairs of nodes to simultaneously communicate at the full bandwidth (either 10 or 100 Mbps) of the network. There are also optical fiber versions of Ethernet and Fast Ethernet. Gigabit Ethernet, providing data transfer rates up to 1 Gbps, has been introduced.

Fiber Distributed Data Interface and Asynchronous Transfer Mode

Fiber distributed data interface (FDDI) is a token-passing network that uses optical fiber cable configured in a ring to transfer information at rates up to 100 Mbps. FDDI can have a second ring transmitting in the opposite direction, permitting the network to function with a break in either ring.

Asynchronous transfer mode (ATM) is a very fast protocol that transmits small packets of uniform size. ATM can be used for LANs and over lines leased from a telecommunications company to connect distant facilities. ATM typically transmits information at 155 Mbps, but data transfer rates can range from 25 to more than 600 Mbps. ATM can use either twisted pair cables carrying electrical signals or optical fiber cables.

Although FDDI or ATM can be used to connect individual nodes on a LAN segment, they are more commonly used as high-capacity data paths called “backbones” between LAN segments in larger and more complex networks. Gigabit Ethernet is also likely to be used for backbones.

Extended Local Area Networks—Repeaters and Bridges

An extended LAN connects facilities, such as the various buildings of a medical center, over a larger area than can be served by a single LAN segment. An extended LAN may be formed by connecting LAN segments with devices such as bridges. Network backbones of high-bandwidth media such as FDDI or ATM may be used to carry heavy information traffic between LAN segments. LANs at distances that cannot be served by the network backbone are usually connected by communications channels provided by a telecommunications company.

Repeaters and bridges can be used to extend or connect LANs. A *repeater* connects two LANs of the same type, thereby extending the distance that can be served. A repeater merely transmits all traffic on each segment of a LAN to the other segment. Because all traffic is repeated, even when the transmitting and destination

nodes are both on the same segment of the network, a repeater does nothing to alleviate network congestion. Repeaters operate at the Physical Layer (Layer 1) of the network protocol stack. A *bridge* is a more sophisticated device that serves the same function as a repeater, to extend a LAN, but transmits packets only when the addresses of the transmitting and destination nodes are on opposite sides of the bridge. Therefore, in addition to extending a LAN, a bridge serves to segment it, thereby reducing congestion. A *Layer 2 switch* can replace the hub at the center of a LAN, permitting multiple simultaneous connections between nodes. Bridges and switches operate at the Data Link Layer (Layer 2) of the network protocol stack.

Very Large Networks and Linking Separate Networks

There are practical limits to the linking of LAN segments by devices such as bridges to form an extended LAN. For example, bridges are not effective at defining the best route across many LAN segments. Furthermore, the addresses of nodes on a LAN, although unique, do not have a hierarchy to assist in directing packets toward their final destination node across very large networks or between separate networks. (The post office would have a similar problem if each postal address consisted of a unique number that did not specify a geographic location.) Additional problems are posed by the heterogeneity of media (e.g., Ethernet, FDDI, ATM) used in individual LAN segments.

Routers

Very large networks and separate networks are usually connected by *routers*. Routers are specialized computers or switches designed to route packets among networks. Routers receive packets and, by following directions in routing tables, forward them on toward their destinations. Each packet may be sent through several routers before reaching its destination. Routers communicate among each other to determine optimal routes for packets. Routers limit network congestion, as do bridges.

Routers follow a routable protocol that assigns each interface in the connected networks a network address distinct from its LAN address. Routers operate at the Network Layer (Layer 3) of the network protocol stack. The dominant routable protocol today is the Internet Protocol.

Transmission Control Protocol/Internet Protocol

The Transmission Control Protocol/Internet Protocol (TCP/IP) is a packet-based suite of protocols used by many large networks and the Internet. TCP/IP permits information to be transmitted from one computer to another across a series of networks connected by routers. TCP/IP operates at protocol layers above those of hardware-layer protocols such as Ethernet and FDDI. The two main protocols of TCP/IP are the *Transmission Control Protocol* (TCP), operating at the Transport Layer (Layer 4) and the *Internet Protocol* (IP), operating at the Network Layer (Layer 3). When an application passes information to TCP/IP, TCP divides the information into packets, adds error-detection information, and passes them to the Network Layer. The Network Layer, following the IP, routes the packets across the networks to the destination computer. Each computer is assigned an *IP address*, consisting of four 1-byte binary numbers, which permits more than 4 bil-

lion distinct addresses. The first part of the address identifies the network, and the latter part identifies the individual computer on the network. (Because of the size of the Internet, efforts are underway to increase the number of addresses permitted by IP.) IP is referred to as a “connectionless protocol” or a “best effort protocol.” That is, the packets are routed across the networks to the destination computer following IP, but some may be lost *en route*. IP does not guarantee delivery or even require verification of delivery. IP also defines how IP addresses are converted to LAN addresses for use by lower-level protocols such as Ethernet. On the other hand, TCP is a connection-oriented protocol that provides reliable delivery. Following TCP, the sending computer initiates a dialog with the destination computer, negotiating matters such as packet size. The destination computer requests the retransmission of any missing or corrupted packets, places the packets in the correct order, recovers the information from the packets, and passes it up to the proper application.

IP addresses, customarily written as four numbers separated by periods (e.g., 152.79.110.12), are inconvenient. Instead, people use host names, such as *web.ucdmc.ucdavis.edu*, to designate a particular computer. The domain name system (DNS) is an Internet service that translates host names into IP addresses.

Internets and the Internet

Networks of networks, using TCP/IP, are called *internets* (with a lower-case “i”). *The Internet* (with a capital letter “I”) is an international network of networks using TCP/IP. A network using TCP/IP within a single company or organization is sometimes called an *intranet*.

Other TCP/IP Protocols

The TCP/IP protocol suite contains several other protocols. These include the file transfer protocol (FTP) for transferring files between computers; simple mail transfer protocol (SMTP) for e-mail; TELNET, which allows a person at a computer to use applications on a remote computer; and hypertext transfer protocol (HTTP), discussed later.

A universal resource locator (URL) is a string of characters used to obtain a service from another computer on the Internet. Usually, the first part of a URL specifies the protocol (e.g., FTP, TELNET, HTTP); the second part of the URL is the host name of the computer from which the resource is requested, and the third part specifies the location of the file on the destination computer. For example, in the URL *http://web.ucdmc.ucdavis.edu/physics/text*, “http” is the protocol, “web.ucdmc.ucdavis.edu,” is the host name, and “physics/text” identifies the location of the information on the host computer.

The World Wide Web (WWW) is the fastest growing part of the Internet. The WWW consists of servers connected to the Internet that store documents, called *web pages*, written in a language called Hypertext Markup Language (HTML). To read a web page, a person must have a computer, connected to the Internet, that contains an application program called a web browser (e.g., Netscape Navigator). Using the web browser, the user types or selects a URL designating the web server and the desired web page. The browser sends a message to the server requesting the web page, and the server sends it, using HTTP. When the browser receives the web page, it is displayed according to the HTML instructions contained in the page. For example, the URL *http://web.ucdmc.ucdavis.edu/*

physics/text will obtain, from a server at the U.C. Davis Medical Center, a web page describing this textbook.

Much of the power of the WWW stems from the fact that a displayed web page may itself contain URLs. If the user desires further information, he or she can select a URL on the page. The URL may point to another web page on the same server, or it may point to a web page on a server on the other side of the world. In either case, the web browser will send a message and receive and display the requested web page. WWW technology is of particular interest in PACS because it can be used to provide images and interpretations to other physicians.

Long-Distance Telecommunications Links

Long-distance telecommunication links are used to connect individual computers to a distant network and to connect distant LANs into a WAN. Most long-distance telecommunication links are provided by telecommunications companies (Table 17-2). In some cases, there is a fixed fee for the link, depending on the maximal bandwidth and distance, regardless of the usage. In other cases, the user pays only for the portion of time that the link is used. Connections to the Internet are usually provided by companies called Internet service providers (ISPs). The interface to an ISP's network can be made by almost any of the telecommunications links described here.

The slowest but least expensive link is by the telephone modem, discussed in Chapter 4. The word *modem* is a contraction of "modulator/demodulator." A telephone modem converts digital information to analog form for transmission over the standard voice-grade telephone system to another modem. As described in Chapter 4, a modem uses frequency modulation, in which the information is encoded in a voltage signal of constant amplitude but varying frequency (tone).

TABLE 17-2. LONG DISTANCE COMMUNICATIONS LINKS PROVIDED BY TELECOMMUNICATION COMPANIES^a

Link	Data transfer rate	Note	Point-to-point?
Modem	28.8, 33.6, and 56 kbps	Analog signals over POTS	No; can connect to any computer or ISP with modem
ISDN	128 kbps	Digital signals over POTS	No
DSL	Various, typically up to DS1.	Digital signals over POTS	Typically provides connection to one ISP
Cable modem	Various, typically 500 kbps to 10 Mbps	Analog signals over a cable television line	Typically provides connection to one ISP
T1	DS1 (1.544 Mbps)	Leased line	Yes
T3	DS3 (44.7 Mbps)	Leased line	Yes
Frame relay	DS1 through DS3	—	No
OC1	51.84 Mbps	Leased line	Yes
OC3	155 Mbps	Leased line	Yes
OC48	2488 Mbps	Leased line	Yes

Note: DSL, digital subscriber line; ISDN, integrated services digital network; ISP, internet service provider; OC, optical carrier; POTS, plain old telephone system.

^aThese can be used to connect an individual computer to a network or to link LANs into a WAN.

Typical data transfer rates of modems are 28.8, 33.6, and 56 kbps. A modem attempts to send at its maximal rate but reduces the rate if the telephone line cannot support the maximal rate. A telephone modem can be used to connect an individual computer to any other modem-equipped computer or network or ISP.

To connect homes and small offices to networks economically while providing higher bandwidth than a modem, some local telephone companies and ISPs employ modalities using the standard copper voice-grade telephone lines already in place. The ISDN and DSL modalities use standard telephone lines to link the home or office to the nearest telephone branch office.

Integrated Services Digital Network (ISDN) permits the transfer of digital signals over several pairs of telephone lines at 128 kbps. A computer using ISDN must be attached to a device called an ISDN terminal adapter, often mistakenly called an "ISDN modem."

Digital subscriber lines (DSL) transfer digital data over local telephone lines at speeds up to about 1.5 Mbps. DSL is usually used to connect computers or networks to the Internet. Some versions of DSL provide higher speeds downstream (to the customer) than upstream; these are referred to as asymmetric digital subscriber lines (ADSL). Telecommunications companies are having varied success with different types of DSL, depending on the length and condition of the local telephone lines.

Cable modems connect computers to the Internet by analog cable television lines. Cable television lines usually support a higher bandwidth than do local telephone lines, often up to or beyond 2 Mbps, providing a high data transfer rate at low cost.

Point-to-point digital telecommunications links may be leased from telecommunications companies. A fixed rate is usually charged, regardless of the amount of usage. These links are available in various capacities (see Table 17-2). The most common are T1, at 1.544 Mbps, and T3, at 44.7 Mbps. Although these links behave as dedicated lines between two points, they are really links across the telecommunications company's network. Optical carrier (OC) links are high-speed links that use optical fiber transmission lines.

Non-point-to-point communications links are also available from telecommunications companies when it is necessary to link several geographically separated facilities into a WAN. Frame relay is a packet-switched technology commonly used for linking multiple facilities into a WAN when information must be transferred between various pairs of the facilities. Frame relay is available in various speeds up to or beyond 1.544 Mbps.

The Internet itself may be used to link geographically-separated LANs into a WAN. Encryption and authentication can be used to create a virtual private network within the Internet.

Network Security

Issues regarding the security of individual computers were discussed in Chapter 4. As mentioned there, the goals of security are to deny unauthorized persons access to confidential information (e.g., patient data) and to protect software and data from accidental or deliberate modification or loss. Computer networks pose significant security challenges, because unauthorized persons can access a computer over the network and because, on many types of LANs, any computer on the network

can be programmed to read the traffic. If a network is connected to the Internet, it is vulnerable to attack by every “hacker” on the planet.

Obviously, the simplest way to protect a network is to not connect it to other networks. However, this may greatly limit its usefulness. For example, at a medical center, it is useful to connect the network supporting a PACS to the hospital information system. If the medical center is affiliated with a university, there may be an interface to the campus network, providing every budding computer scientist on the campus with a pathway into the PACS. An interface to the campus network provides a link to the Internet as well.

In such cases, a *firewall* can enhance the security of a network. A firewall is a router, a computer, or even a small network containing routers and computers that is used to connect two networks to provide security. There are many different services that can be provided by a firewall. One of the simplest is packet filtering, in which the firewall examines packets reaching it. It reads their source and destination addresses and the applications for which they are intended, and, based on rules established by the network administrator, forwards or discards them. For example, packet filtering can limit which computers a person outside the firewall can access and can forbid certain kinds of access, such as FTP and TELNET. Different rules may be applied to incoming and outgoing packets. In this way, the privileges of users outside the firewall can be restricted without limiting the ability of users on the protected network to access computers past the firewall. Firewalls also can maintain records of the traffic across the firewall, to help detect and diagnose attacks on the network. A firewall, by itself, does not provide complete protection, and therefore it should merely be part of a comprehensive computer security program. In particular, a firewall does not eliminate the need for the individual computer security measures described in Chapter 4.

Three important security issues regarding networks, particularly when information is sent between distant sites, are privacy, authentication, and integrity. Privacy, also called confidentiality, refers to preventing persons other than the intended recipient from reading the transmitted information. Encryption maintains the privacy of a message by translating the information into a code that only the intended recipient can convert to its original form. Authentication permits the recipient of information to verify the identity of the sender and permits the sender to verify the identity of the recipient. Integrity means that the received information has not been altered, either deliberately or accidentally. There are methods, beyond the scope of this text, that permit the recipient and sender to authenticate each other and that permit the recipient to verify the integrity of a message.

17.2 PACS AND TELERADIOLOGY

A PACS is a system for the storage and transfer of radiologic images. Teleradiology is the transmission of such images for viewing at a site or sites remote from where they are acquired. PACS and teleradiology are not mutually exclusive. Many PACS incorporate teleradiology. It is essential that PACS and teleradiology provide images suitable for the task of the viewer. In addition, when the viewer is the interpreting physician, the images viewed must not be significantly degraded in either contrast or spatial resolution with regard to the acquired images.

Picture Archiving and Communications Systems

A PACS consists of interfaces to imaging devices that produce digital images, possibly devices to digitize film images, a digital archive to store the images, display workstations to permit physicians to view the images, and a computer network to link them. There also must be a database program to track the locations of images and user software that allows interpreting physicians to select and manipulate images. Links to other networks, such as the hospital information system (HIS) and radiology information system (RIS), are useful.

PACS vary widely in size and in scope. A PACS may be devoted to only a single modality at a medical facility, such as a nuclear medicine department, the ultrasound section, a couple of cardiac catheterization laboratories, or the x-ray computed tomography (CT) and magnetic resonance imaging (MRI) facilities. In this case, the entire PACS may be connected by a single LAN. Such a small, single-modality PACS is sometimes called a *mini-PACS* (Fig. 17-3). On the other hand, a PACS may incorporate all imaging modalities in a system of several medical centers and affiliated clinics (Fig. 17-4). In addition, a PACS may permit images to be viewed only by interpreting physicians or it may make them available to the emergency department, intensive care units, and attending physicians as well. In these cases, the PACS is likely to exist on an extended LAN or on a WAN constructed from multiple LANs.

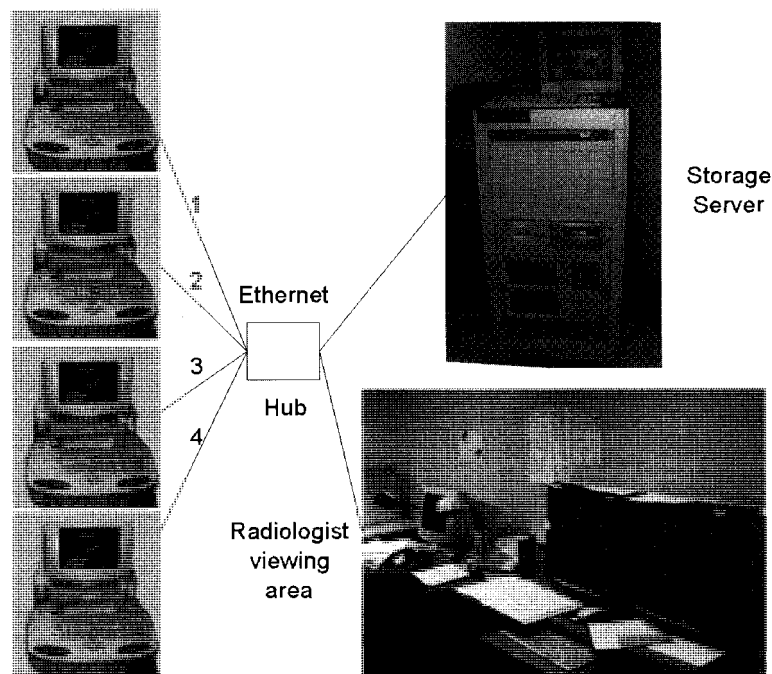


FIGURE 17-3. Mini-PACS. A mini-PACS is a picture archiving and communications system serving a single modality, such as ultrasound, nuclear imaging, or computed tomography. Shown is an ultrasound mini-PACS that serves four ultrasound scanners and is connected by a hub to a file server and several radiologist workstations.

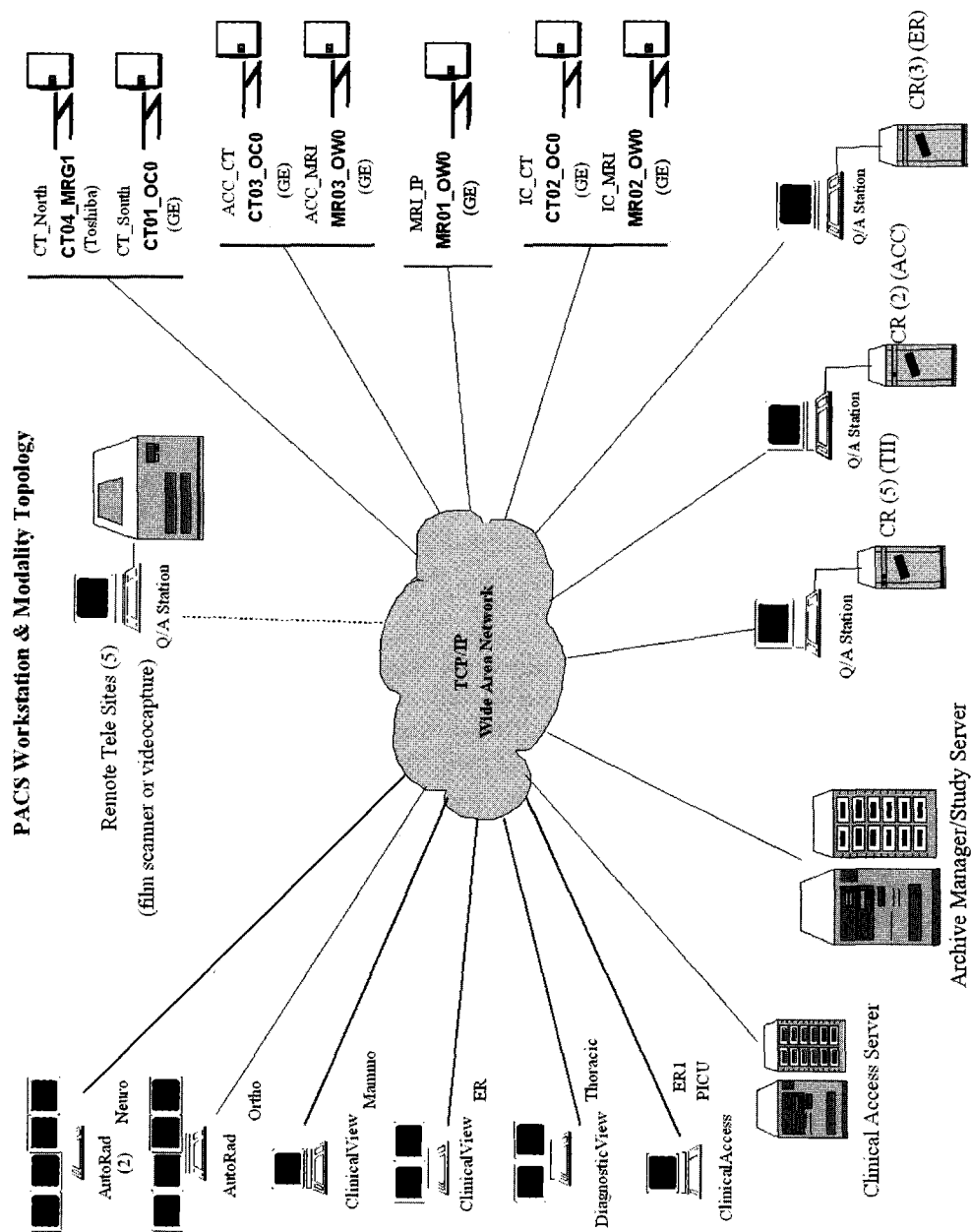


FIGURE 17-4. A picture archiving and communications system (PACS) serving a large medical system.

Teleradiology

Teleradiology can provide improved access to radiology for small medical facilities and improved access to specialty radiologists. For example, teleradiology can permit radiologists at a medical center to promptly interpret radiographs acquired at affiliated outpatient clinics. This requires film digitizers or devices producing digital radiographs at each clinic, leased lines to the medical center, and an interpretation workstation at the medical center. In another example, teleradiology can permit an on-call radiologist to interpret after-hours studies at home, avoiding journeys to the medical center and providing interpretations more promptly. This requires the radiologist to have an interpretation workstation at home, perhaps with an ISDN or DSL connection provided by a local telecommunications company, and the medical center to have images in a digital format and a matching communications connection.

Acquisition of Digital Images

In many modalities, the imaging devices themselves produce digital images. CT, MRI, digital radiography, digital fluoroscopy, ultrasound, and nuclear medicine images are already in digital formats. However, large fractions of radiographic images are still acquired with the use of film-screen image receptors. Film images can be digitized by laser or charge-coupled device (CCD) digitizers.

Image Formats

Image formats are selected to minimize the degradation of the images. Therefore, imaging modalities that provide high spatial resolution (e.g., radiography) require large pixel formats, and modalities that provide high contrast resolution (e.g., x-ray CT) require a large number of bits per pixel. Typical formats for digital images were listed in Table 4.7 of Chapter 4.

The American College of Radiology (ACR) has published standards for teleradiology and for digital image data management. These standards categorize images with small or large matrices and provide specifications regarding each. They specify that large matrix images (computed radiography and digitized radiographic films) should allow a spatial resolution of at least 2.5 line pairs (lp) per millimeter at the original detector plane and should have at least 10 bits per pixel.

Consider, for example, a chest radiograph (35 by 43 cm) and apply the ACR criterion for large-matrix images:

$$\text{Vertical lines} = (35 \text{ cm}) (2.5 \text{ lp/mm}) (2 \text{ lines/lp}) = 1,750 \text{ lines}$$

$$\text{Horizontal lines} = (43 \text{ cm}) (2.5 \text{ lp/mm}) (2 \text{ lines/lp}) = 2,150 \text{ lines}$$

Thus, the ACR's spatial resolution standard requires a pixel format of at least $1,750 \times 2,150$ for such images.

Film Digitizers

A film digitizer is a device that scans the film with a light source, measures the transmitted light, and forms a digital image depicting the distribution of optical density (OD) in the film. There are two types of digitizers commonly used, one employing

a laser and the other using a collimated light source and a CCD linear array. In general, the laser digitizer measures OD more accurately at high ODs, whereas the CCD digitizer is less costly and requires less maintenance.

In a laser digitizer, the film is scanned one row at a time by a laser beam. The intensity of the beam transmitted through the film is measured by a light detection system that incorporates a photomultiplier tube (PMT) or photodiode. The signal from the light detector is logarithmically amplified and then digitized by an analog-to-digital converter (ADC) (Fig. 17-5). Logarithmic amplification produces a signal that corresponds linearly to film OD. (Recall from Chapter 6 that the OD is $\log_{10} (1/T)$, where T is the fraction of incident light that is transmitted through the film.)

In a CCD digitizer, the film is continuously moved across a uniformly lighted slit. A one-dimensional CCD array converts the transmitted light into an electrical signal, one row at a time (Fig. 17-6) (CCDs are discussed in Chapter 11.) This is a much simpler design than the laser digitizer. In some CCD designs, the signal from the CCD is logarithmically amplified before digitization. Alternatively, the logarithm of the linearly digitized signal can be obtained with a logarithmic look-up table.

Light transmitted through the film varies exponentially as a function of OD. Logarithmic amplification of the output signal produces a linear relationship between the digital number and the OD. This can be done before digitization with an analog logarithmic amplifier. If the logarithmic operation is applied after digitization, there is a loss of accuracy because fewer digital values are available at high OD. With a 12-bit ADC (4,096 gray levels), there are fewer than 400 digital val-

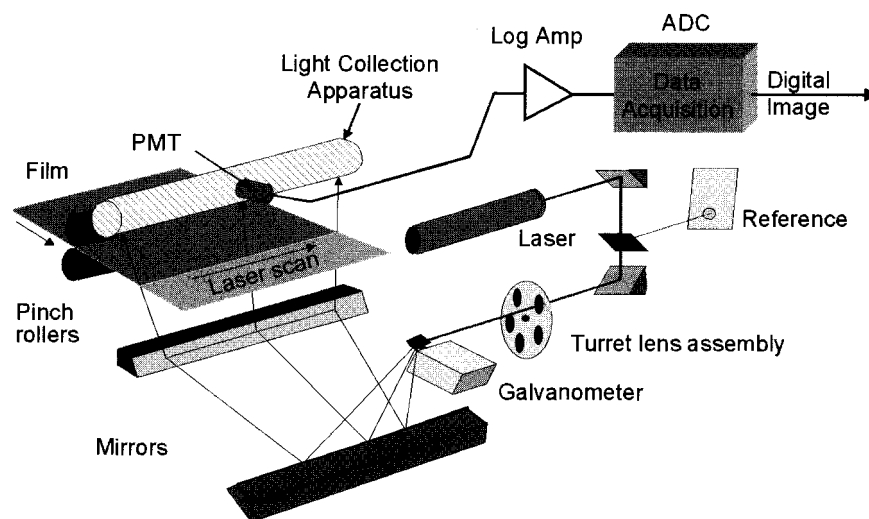


FIGURE 17-5. Components of a laser film digitizer. A developed film is transported past the light collection apparatus. The laser beam is selectively focused by an optical lens on the turret lens assembly and is reflected by a mirror attached to a galvanometer that tilts back and forth, creating a laser scan pattern. The light collection apparatus is a reflective tube that permits the laser beam transmitted through the film to be measured. The light entering the tube through a slit reaches the photomultiplier tube (PMT) by multiple scattering. Efficiency of light collection varies with distance from the PMT (e.g., fall-off of intensity occurs at the periphery of the scan), which is corrected by digital methods.

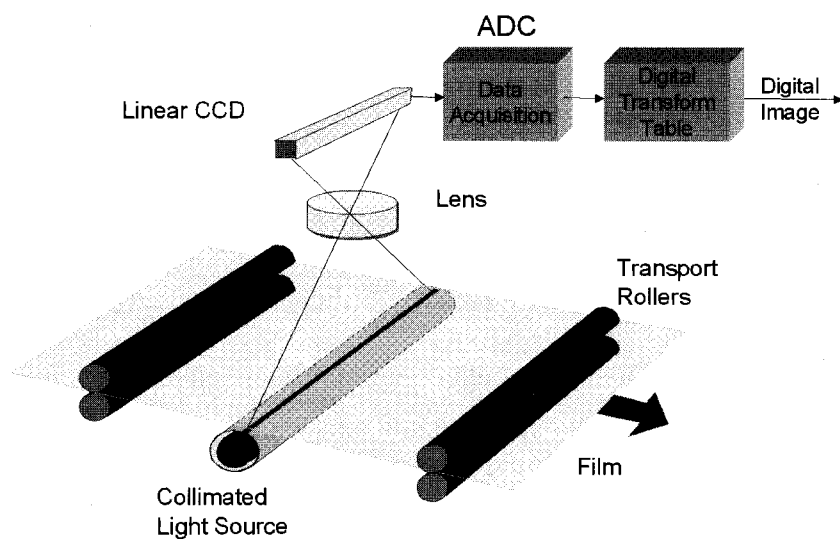


FIGURE 17-6. Charge-coupled device (CCD) film digitizer. The typical light source is a standard fluorescent tube. The linear CCD photosensor is composed of up to 4,000 pixels in a single row. Each pixel of the CCD produces electrons in proportion to the light transmitted through at a corresponding location on the film. The CCD pixel values are read one at a time, digitized, and formed into a digital image.

ues representing ODs above 1.0, and fewer than 40 digital values for ODs greater than 2.0 (Fig. 17-7). A change of ± 0.1 OD is represented by only about 4 digital values in this range.

The spatial resolution produced by film digitizers should meet the ACR standards and be similar to the equivalent digital modality. For instance, a 35×43 cm

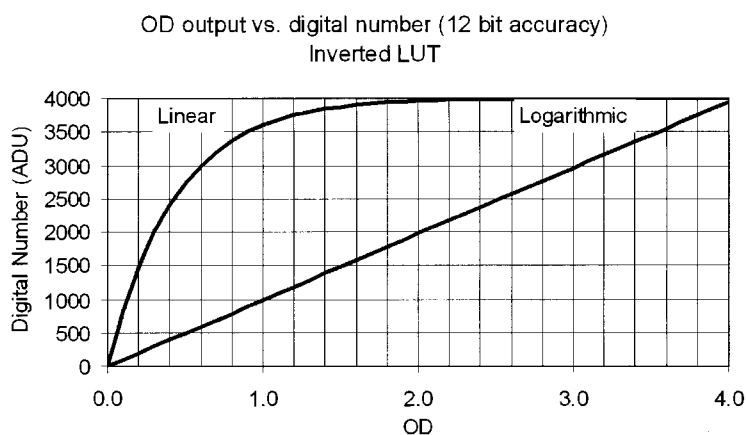


FIGURE 17-7. Digital values from the analog-to-digital converter (ADC) for linear and logarithmic amplification of the input signal to the ADC, as a function of film optical density (OD). (The signals are inverted so that the digital number is proportional to OD.) If logarithmic amplification is not applied before digitization, large changes in OD at high ODs produce only tiny changes in the digital numbers. LUT, look-up table.

(14 × 17 inch) film should be digitized to a pixel size of at least 200 μm (1,750 × 2,150 pixel matrix). Most CCD and laser digitizers meet these requirements.

The role of film digitizers in PACS is likely to be transient because of the decreasing use of screen-film detectors. However, it is likely to be a decade or longer before film's role in recording and displaying images will be but a footnote in a future edition of this book.

Digital Imaging and Communications in Medicine

In the past, imaging equipment vendors have used proprietary formats for patient data and digital images, thereby hindering the transfer, storage, and display of images acquired on equipment from various vendors and making it difficult to build PACS and teleradiology systems. Some facilities solved this problem by purchasing all equipment from a single vendor. To help overcome these problems, the ACR and the National Electrical Manufacturers' Association (NEMA) jointly sponsor a standard called Digital Imaging and Communications in Medicine (*DICOM*) to facilitate the transmission of medical images and related data. DICOM specifies standard formats for information objects, such as "patients," "studies," and "images." DICOM also specifies standard services that may be requested regarding these information objects, such as "find," "move," "store," and "get." Usually DICOM is used as an upper-layer protocol on top of TCP/IP.

Today, the products of most major vendors of medical imaging and PACS equipment conform to the DICOM standard. A *DICOM conformance statement* is a formal statement, provided by a vendor, that describes a specific implementation of the DICOM standard. It specifies the services, information objects, and communications protocols supported by the implementation. In reality, even when the conformance statements issued by two vendors indicate the ability to provide a particular service, interoperability is not assured. Vendors should be contractually obligated to provide necessary functions, and verification testing must be performed after installation.

Networks for Image and Data Transfer

Computer networks, discussed in detail at the beginning of this chapter, are used to link the components of a PACS. A PACS may have its own network, or it may share network components with other information systems. The network may provide images only to interpretation workstations, or it also may provide them to display stations in intensive care units and in the emergency department. The bandwidth requirements depend on the imaging modalities and their work loads. For example, a network medium adequate for nuclear medicine or ultrasound may not be adequate for other imaging modalities. Network traffic typically varies in a cyclic fashion throughout the day. Network traffic also tends to be "bursty." There may be short periods of very high traffic, separated by long periods of low traffic. Network design must take into account both peak and average bandwidth requirements and the delays that are tolerable over each segment of the network. Network segmentation, whereby groups of imaging, archival, and display devices that communicate frequently with each other are placed on separate segments, is commonly used to reduce network congestion. Different network media may be used for each network segment. For example, 10Base-T Ethernet is likely to be sufficient for a segment

serving nuclear medicine, whereas a network segment carrying CT, MRI, and digital radiographs may require a higher-bandwidth medium such as Fast Ethernet or FDDI.

Interfaces to the Radiology Information System and Hospital Information System

It is advantageous to have interfaces between a PACS, the RIS, and the HIS. A RIS is used for functions such as ordering and scheduling procedures, maintaining a patient database, transcription, reporting, and bill preparation. The RIS is not always a separate system—it may be part of the HIS or incorporated in the PACS. Such interfaces can provide worklists of requested studies, thereby providing patient data to the PACS and reducing the amount of manual data entry. The PACS or RIS should supply the operator's consoles of the imaging devices with worklists of patient descriptors to identify each study. Communication among the RIS, the HIS, and the PACS is often implemented using a standard called Health Level 7 (HL7) for the electronic exchange of medical information (e.g., administrative information, clinical laboratory data). HL7 to DICOM translation, for instance, provides worklists to the imaging devices from the RIS or HIS patient database.

Using patient identifying information supplied by the RIS or HIS reduces a common problem in PACS—the inadvertent assignment of different patient identifiers to studies of a specific patient. This can arise from errors in manual data entry, such as incorrect entering of a Social Security number by a technologist or clerk. Entering a patient's name in different formats—such as Samuel L. Jones, S. L. Jones, and Jones, Samuel—also may cause problems. These problems are largely obviated if the technologists must select only patient identifiers from worklists, provided of course that the RIS and/or HIS furnishes unique patient identifiers.

These worklists also permit prefetching of relevant previous studies for comparison by interpreting physicians. The network interfaces to the RIS and HIS can provide the interpreting physicians with interpretations of previous studies and the electronic patient record via the PACS workstation, saving the time that would be required to obtain this information on separate computer terminals directly connected to the RIS and/or HIS. These interfaces can also permit the PACS to send information to the RIS and/or HIS regarding the status of studies (e.g., study completed).

Provision of Images and Reports to Attending Physicians and Other Health Care Providers

WWW technology provides a cost-effective means for the distribution of images and other data over the HIS to noninterpreting physicians and other health care professionals. A web server, interfaced to the PACS, RIS, and HIS, maintains information about patient demographics, reports, and images. Data is sent from the PACS, typically in a study summary form (i.e., the full image data set is reduced to only those images that are pertinent to a diagnosis). The full image set may be available as well. The server provides full connectivity to the PACS database to allow queries and retrieval of information when requested. Physicians obtain these images and reports from the web server using workstations with commercial web browsers. A major advantage to using WWW technology is that the workstations need not be

equipped with specialized software for image display and manipulation; instead, these programs can be sent from the web server with the images.

Healthcare professionals at remote sites (e.g., doctors' offices, small clinics, at home) can connect to the HIS using modems or ISDN or DSL lines. Alternatively, if the web server is interfaced to the Internet, these professionals can obtain images and reports over the Internet.

Storage of Images

In a PACS, a hierarchical storage scheme has been commonly used, with recent images available on arrays of magnetic hard disks and with a culling and transfer of older images to slower but more capacious archival storage media, such as optical disks and magnetic tape. The amount of storage necessary depends on the modality or modalities served by the PACS and on the work load. For example, a medium-sized nuclear medicine department generates a few gigabytes (uncompressed) of image data in a year, so a single, large-capacity magnetic disk and a pair of WORM or rewriteable optical disk drives would suffice for a mini-PACS serving such a department. On the other hand, CT, MRI, and digitized radiographs from a medium-size radiology department can generate several gigabytes of image data in a day and several terabytes (TB) (uncompressed) in a year, requiring a much larger and more complex archival system. Ultrasound storage requirements strongly depend on the number of images in each study that are selected for archiving. For a medium-sized medical center, if 60 images per patient are selected for storage, then up to 0.5 TB per year may be generated. A cardiac catheterization laboratory may generate up to a few terabytes (uncompressed) in a year.

The term *on-line storage* is used to describe storage, typically arrays of magnetic disks, that is used to provide almost immediate access to studies. The term *near-line storage* refers to storage, such as automated arrays ("jukeboxes") of optical disks or magnetic tape, from which studies may be retrieved within 1 minute without human intervention (Fig. 17-8). Optical disks have a slight advantage over magnetic tape cassettes for near-line storage because of shorter access times, but magnetic tape is less expensive. *Off-line storage* refers to optical disks or magnetic tape cassettes stored on shelves or racks. Obtaining a study from off-line storage requires a person to locate the relevant disk or cassette and to manually load it into a drive.

A technology called *RAID* (redundant array of independent disks) can be used to provide a large amount of on-line storage. RAID permits several or many small and inexpensive hard magnetic disk drives to be linked together to function as a single very large drive (Fig. 17-9). There are several implementations of RAID, designated as RAID Levels 0 through 5. In RAID Level 0, portions of each file stored are written simultaneously on several disks, with different portions on each disk. This produces very fast read and write capability but with no redundancy, so the loss of one of these drives results in the loss of the file. In RAID Level 1, called *disk mirroring*, all information is written simultaneously onto two or more drives. Because duplicate copies of each file are stored on each of the mirrored disks, Level 1 provides excellent data protection but without any improvement in data transfer rate and with greatly increased storage cost. The other RAID levels provide various compromises among fault tolerance, data transfer rate, and storage capacity.

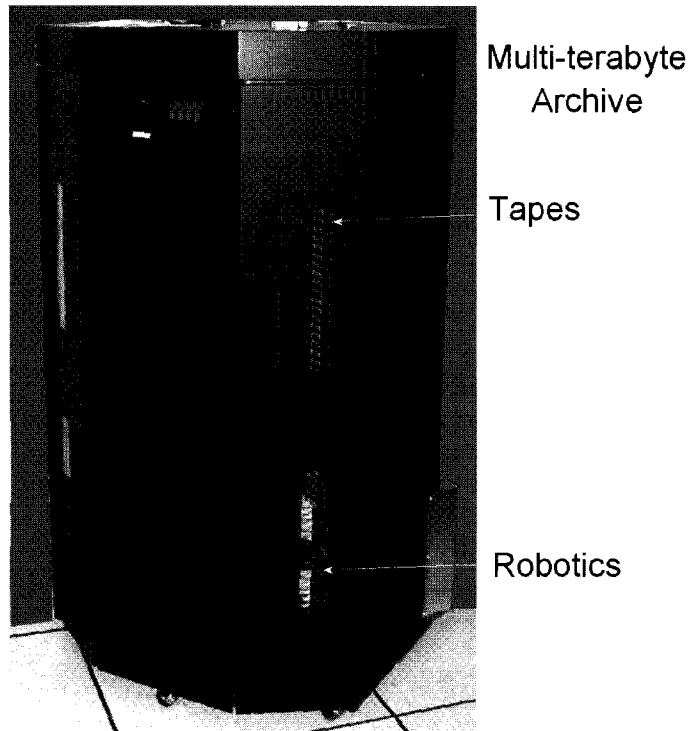


FIGURE 17-8. Robotic digital linear tape “jukebox.” This jukebox has two tape drives and permits storage of 1,500 0.75-GB tape cassettes, for a total storage capacity of approximately 1 TB, not compressed.

$$14 \times 36.4 \text{ GB} = 500 \text{ GB}$$

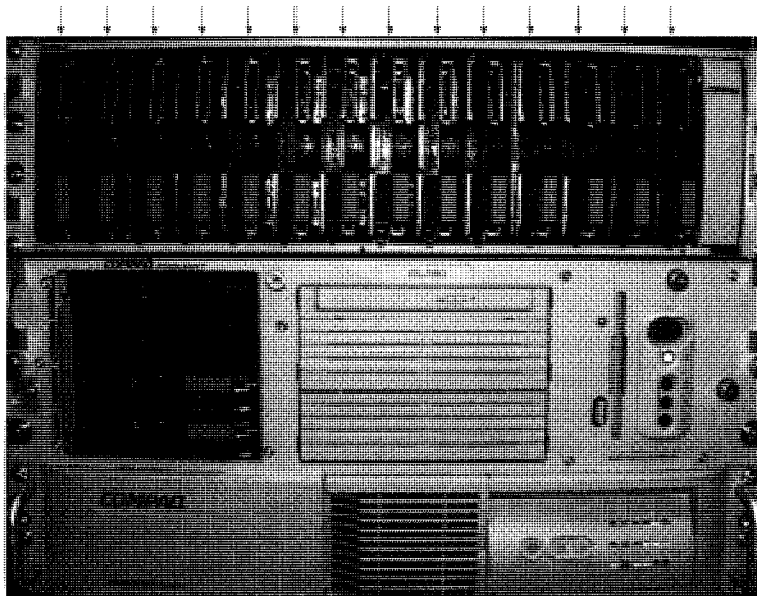


FIGURE 17-9. Redundant array of independent disks (RAID). This array contains fourteen 36.4-GB magnetic hard disk drives, with a total active storage capacity of 500 GB. (Some disk space is used for RAID fault tolerance, making the total available uncompressed storage capacity substantially less.)

The storage may be centralized (i.e., all attached to a single computer, called a storage server), or it may be distributed (i.e., stored at several locations around a network). In either case, there must be a database program to track the studies and their locations. There must also be image management software to perform prefetching algorithms, image transfer to various levels of archive storage, and deletion from local storage on workstations.

Image Management

In some PACS systems, studies awaiting interpretation and relevant older studies are requested from the archive as needed ("on demand") during interpretation. Alternatively, they may be obtained from the archive and stored on the local magnetic disks of the interpretation workstation, before the interpretation session, ready for the interpreting physician's use ("prefetching"). The latter method requires the interpretation workstations to have more local storage capacity, whereas the former requires a faster archive and faster network connections between the archive and the interpretation workstations. When images are fetched on demand, the first image should be available for viewing within about 2 seconds. The best strategy is most likely a hybrid system that functions, when necessary, as an on-demand system but also provides prefetching of images, when possible, to reduce network congestion during peak traffic periods. Once the images are residing on the local workstation disk, they are almost instantaneously available.

Data Compression

The massive amount of data in radiologic images poses considerable challenges regarding storage and transmission (Table 17-3). Image compression reduces the number of bytes in an image or set of images, thereby decreasing the time required to transfer images and increasing the number of images that can be stored on a magnetic disk or unit of removable storage media. Compression can reduce costs by permitting the use of network links of lower bandwidth and by reducing the amount of required storage capacity. Before display, the image must be decompressed. Image compression and decompression can be performed either by a general-purpose

TABLE 17-3. IMAGE STORAGE REQUIREMENTS FOR VARIOUS RADIOLOGIC STUDIES, IN UNCOMPRESSED FORM

Study	Typical storage requirement (megabytes)
Chest radiographs (posteroanterior and lateral, 2×2 k)	20
Typical computed tomographic study (120 images, 512×512)	64
Thallium 201 myocardial perfusion SPECT study	1
Ultrasound (60 images to PACS archive, 512×512)	16
Cardiac catheterization laboratory study (coronary and left ventricular images)	450–3,000
Digital screening mammograms (2 CC and 2 MLO)	32–220

Note: PACS, picture archiving and communications systems; SPECT, single photon emission computed tomography.

computer or by specialized hardware. Image compression and decompression are calculation-intensive tasks and can delay image display.

There are two categories of compression: lossless and nonrecoverable (lossy) compression. In lossless compression, the uncompressed image is identical to the original. Typically, lossless compression of medical images permits approximately 3:1 compression. Lossless compression takes advantage of redundancies in data. It is not possible to store random and equally likely bit patterns in less space without the loss of information. However, medical images incorporate considerable redundancies, permitting them to be converted into a more compact representation without loss of information. For example, although an image may have a dynamic range (difference between maximal and minimal pixel values) that requires 12 bits per pixel, values usually change only slightly from pixel to pixel, and so changes from one pixel to the next can be represented by just a few bits. In this case, the image could be compressed without a loss of information by storing the differences between adjacent pixel values instead of the pixel values themselves.

In nonrecoverable (lossy) compression, information is lost and the uncompressed image will not exactly match the original image. However, lossy compression permits much higher compression rates; ratios of 15:1 or higher are possible with very little loss of image quality. Currently, there is controversy regarding how much compression can be tolerated. Research shows that the amount of compression depends strongly on the type of examination, the compression algorithm used, and the way that the image is displayed (e.g., film, video display). In some cases, images that are lossy-compressed and subsequently decompressed are actually preferred by radiologists over the original images, because of the selective reduction of image noise. Legal considerations also affect decisions on the use of lossy compression in medical imaging.

Example: How much time would be required to transmit a pair of chest images (2,048 × 2,048 pixels, 2 bytes per pixel) to a radiologist's home using ISDN (128 kb/sec)?

$$\frac{2 \times 2,048 \times 2,048 \text{ pixels} \times 16 \text{ bits/pixel}}{128 \text{ kb/sec} \times 1000 \text{ bits/kb}} = 1,049 \text{ sec} = 17.5 \text{ min}$$

With 15:1 compression, the transmission time would be reduced ideally to 1.2 minutes.

Display of Images

Images from a PACS may be displayed on electronic devices, such as cathode ray tube (CRT) or flat panel video displays. They may also be recorded by a video or laser camera, on photographic film that is chemically developed, and then viewed on a lightbox. In all cases, the information in the digital image is converted to analog form using digital to analog conversion (DACs) for display.

Display Workstations

Computer workstations equipped with video monitors are increasingly used, instead of viewboxes and photographic film, to display radiologic images for both interpretation and review (Fig. 17-10). The various radiologic modalities impose specific requirements on workstations. For example, a workstation that is suitable for the interpretation of nuclear medicine or ultrasound images may not be adequate for digitized radiographs. Workstations for the interpretation of nuclear med-

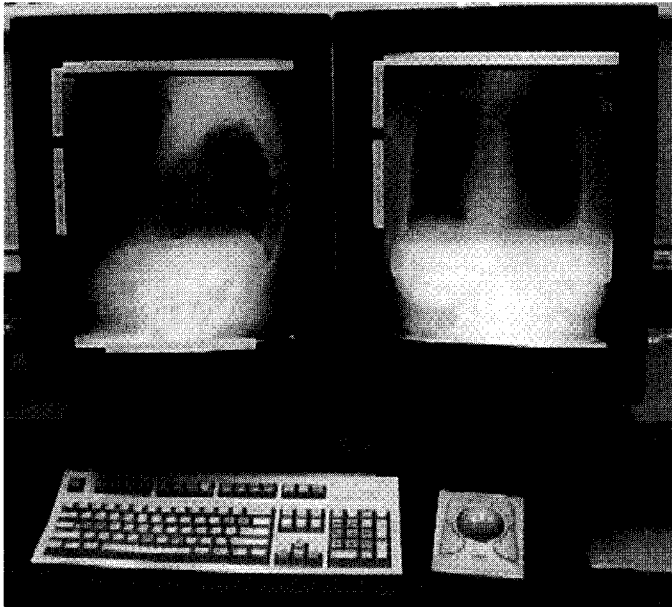


FIGURE 17-10. Interpretation workstation containing two 2×2.5 -k pixel, portrait-format, gray-scale, cathode ray tube (CRT) monitors.

icine and ultrasound images should employ color monitors and must be able to display cine images. Workstations used for interpretation of angiography image sequences (e.g., cardiac angiograms and ventriculograms, digital subtraction angiography) should also be capable of displaying cine images. Modalities other than nuclear medicine and ultrasound usually use gray-scale monitors.

Today, most monitors used for radiologic image display are based on CRTs, but the use of flat panel displays is increasing. High-quality flat panel monitors today use backlit, active-matrix liquid crystal display (LCD) technology. The term *flat panel*, when used throughout this section, refers to this technology. The principles of these devices were described in Chapter 4, and only their performance characteristics are discussed here. Video monitors are part of the imaging chain and are characterized by parameters such as spatial resolution, spatial distortion, contrast resolution, aspect ratio, maximal luminance, dynamic range, uniformity of luminance, noise, persistence, and refresh rate.

The quantity describing the brightness of a monitor (or other light source) is *luminance*. Luminance, defined in Chapter 8, is a linear quantity (i.e., twice the rate of light emission doubles the luminance) that takes into account the varying spectral sensitivity of the human eye to light. The contrast resolution of a monitor is mainly determined by its dynamic range, defined as the difference between its maximal and minimal luminance. (Sometimes, the dynamic range is defined as the ratio of the maximal to minimal luminance.) Especially bright monitors should be used for the display of radiologic images to provide adequate dynamic range. The current ACR standards for teleradiology and digital image data management specify that gray-scale monitors of interpretation workstations should provide a luminance of at least 160 candela (cd) per square meter. Even so, the maximal luminance of a high-performance monitor is less than that of a bright viewbox by about a factor of ten (Table 17-4). In general, gray-scale monitors, whether using CRTs or flat panel

TABLE 17-4. CHARACTERISTICS OF DISPLAYS AND FILM ON VIEWBOXES^a

Image display	Maximal luminance (cd/m ²)
Standard viewbox	1,500–3,500
Mammography viewbox	3,000–7,500 (ACR criterion: $\geq 3,000$)
Grayscale monitor (CRT or flat panel)	120–800 (ACR standard for monitors used for interpretation: ≥ 160)
Color monitor (CRT or flat panel)	80–300
Standard personal computer color monitor	65–130

Note: ACR, American College of Radiology; CRT, cathode ray tube.

^aInformation on viewbox luminance from *Mammography Quality Control Manual*. Reston, VA: American College of Radiology, 1999:286–294. Note: 1 cd/m² = 0.292 foot-lamberts.

displays, provide much higher maximal luminance than do color monitors. CRT monitors of large matrix formats provide lower maximal luminance than do CRT monitors of smaller matrix formats, but the maximal luminance of a flat panel monitor is not significantly affected by the matrix format. The luminance of a CRT monitor is highest at the center and diminishes toward the periphery. The minimal luminance of a monitor is called the *black level*, and CRT monitors have lower measured black levels than flat panel monitors do. In practice, the minimal luminance is limited by veiling glare, the scattering of light inside the monitor. The amount of veiling glare is determined by the size and brightness of other areas of the displayed image and the distances between bright and dark areas. Flat panel monitors suffer much less from veiling glare than do CRT monitors. The dynamic range of a monitor is much less than that of radiographic film on a viewbox, and windowing and leveling are needed to compensate.

Manufacturers often describe spatial resolution in terms of a monitor's addressable pixel format (e.g., $1,280 \times 1,024$ pixels), but this is only a crude indication of spatial resolution. In fact, two monitors with the same pixel format may have quite different spatial resolutions. The spatial resolution of a CRT video monitor, in the direction perpendicular to the scan lines, is determined mainly by the number of scan lines and the width, in the perpendicular direction, of the electron beam striking the phosphor. The profile of the electron beam and thus the point-spread function (PSF) of the light emitted by the phosphor are approximately Gaussian in shape. The full width at half-maximum of the PSF must be comparable to the width of a scan line to reduce the visibility of the scan lines. This causes a significant fraction of the energy in the beam to be deposited in the two adjacent scan lines, degrading the spatial resolution. Simply put, a CRT monitor cannot produce a completely black scan line immediately adjacent to a bright one.

The spatial resolution in the direction parallel to the scan lines is determined mainly by the bandwidth of the video circuitry and the width of the electron beam in that direction. The bandwidth determines how quickly the electron beam intensity can be changed in relation to the time required to scan each line. Spatial resolution often becomes worse toward the periphery of the CRT. A flat panel monitor using active-matrix LCD technology provides spatial resolution superior to that of a CRT monitor with the same addressable pixel format, because a bright pixel has less effect on its neighbors.

Spatial linearity (freedom from spatial distortion) describes how accurately shapes and lines are presented on the monitor. The refresh rate is the number of times per second that the electron beam scans the entire face of the CRT. It should exceed the flicker fusion frequency of the eye (about 50 cycles/sec) so that the human observer does not perceive flicker in stationary images and perceives continuous motion when viewing cine images. Monitor refresh rates typically range from 60 to 120 Hz (images per second). Increasing the refresh rate eliminates the perception of flicker; however, the bandwidth increases proportionately, placing greater demands on the display electronics. Persistence is a delayed emission of light from the screen phosphor that continues after the screen is refreshed. Persistence can be desirable when viewing static images, because it reduces flicker and statistical noise. Excessive persistence can produce lag in a cine display of images; on the other hand, persistence reduces the apparent noise of the image sequence by signal averaging. Monitors add both spatial and temporal noise to displayed images. In CRT monitors, phosphor granularity is a major source of spatial noise, whereas fluctuations in the electron beam intensity contribute to the temporal noise.

Standards for the performance of video monitors for medical image display are currently being developed. Given the rapidly increasing use of video monitors for the interpretation of images of high spatial resolution, particularly digitized radiographs, such standards and standardized test methods are urgently needed.

Monitor Pixel Formats

Monitors are available in several pixel formats. Two common formats for color monitors are 1,280 horizontal pixels by 1,024 vertical pixels and $1,600 \times 1,200$ pixels. A common format for gray-scale monitors used for the display of digitized radiographs is 2,048 horizontal pixels by 2,560 vertical pixels, called a $2\text{-} \times 2.5\text{-k}$ display, or a 5-megapixel display. A monitor and display oriented so that the horizontal field is greater than the vertical is often called a *landscape display*, whereas an orientation with the vertical field greater than the horizontal is called a *portrait display*. Monitors with pixel formats of at least $1,024 \times 1,024$, but less than $2,048 \times 2,048$, are commonly given the generic term "1-k monitors"; those with formats of $2,048 \times 2,048$ or a little more are called "2-k monitors."

Large-matrix images (digitized radiographs) are typically stored in formats of about $2 \times 2\text{ k}$ or $2 \times 2.5\text{ k}$. The $2\text{-} \times 2.5\text{-k}$ "portrait" format permits the display of an entire large matrix image at near-maximal resolution. Even so, some loss of spatial resolution occurs for the reasons described previously. A less-expensive alternative is to use 1-k monitors. However, only portions of the image are displayed when they are viewed in maximal resolution on these monitors. The "pan" function on the workstation allows the viewing of different portions of the image. For the entire image to be displayed, it must be reduced to a smaller pixel format by averaging of neighboring pixels in the stored image.

An interpretation workstation for large-matrix images is usually equipped with at least two monitors to permit the simultaneous comparison of two images in full fidelity. A single 1-k color monitor is sufficient for a nuclear medicine or ultrasound workstation. Noninterpreting physicians typically use display stations each with a single 1-k monitor.

An application program on the workstation permits the interpreting physician to select images, arrange them for display, and manipulate them. The way in which the program arranges the images for presentation and display is called a *hanging protocol*. The hanging protocol should be configurable to the preferences of individual physicians. Image manipulation capabilities of the program should include window and level, zoom (magnify), and roam (pan). The program should also provide quantitative tools to permit accurate measurements of distance and the display of individual pixel values in modality-appropriate units (e.g., CT numbers). Image processing, such as smoothing and edge enhancement, may be provided.

Calibration of Display Workstations

The display function of a video monitor describes the luminance produced as a function of either the digital input value or the magnitude of the analog video signal. The display function of a monitor greatly affects the perceived contrast of the displayed image. The display functions of CRT monitors are nonlinear (i.e., increasing the input signal does not, in general, cause a proportional increase in the luminance) and vary from monitor to monitor.

A CRT monitor has controls labeled "brightness" and "contrast" that are used to adjust the shape of the display function. On most monitors, these controls are not equipped with scales or detents so that they can be set reproducibly; therefore, once they are adjusted, they should not be changed by the user. On most monitors used for image interpretation, these controls are not available to the user.

The display function of a monitor may be modified to any desired shape by the use of look-up tables in the workstation (see Chapter 4). The look-up table contains an output pixel value for each possible pixel value in an image. By establishing a look-up table for each monitor, all of the monitors at a single interpretation workstation or in a radiology department can be calibrated to provide similar display functions. This calibration can be performed manually, but it is best performed automatically by the workstation itself, using a photometer, aimed at a single point on the face of the monitor, whose output is digitized and routed into the workstation. Similarly, additional look-up tables can be used to correct the nonuniformity of luminance across the face of the monitor.

The optimal display function may vary with the imaging modality. Furthermore, it should take into account previous steps in the imaging chain. For example, a film digitizer can produce images with pixels representing either transmittance (T) or optical density (OD), where $OD = \log_{10} (1/T)$. The optimal display function is likely to be different in these two cases. Efforts are in progress to determine optimal display functions. The ability to window and level, selectively enhancing the contrast within any range of pixel values, reduces the importance of optimizing the display function.

Ambient Viewing Conditions

The ambient viewing conditions at the sites where the interpretation workstations are used are important to permit the interpreting physicians' eyes to adapt to the low luminance of the monitors, to avoid a loss in contrast from diffuse glare from

the faces of the monitors, and to avoid reflections on the monitor's faces from bright objects. Viewing conditions are more important when monitors are used than when viewboxes are used, because of the lower luminance of the monitors. The room should have adjustable indirect lighting and should not have windows unless they can be blocked with opaque curtains or shutters. When more than one workstation is in a room, provisions should be made, either by monitor placement or by the use of partitions, to prevent the monitors from casting reflections on each other. Particular care must be taken, if both conventional viewboxes and workstations are used in the same room, to prevent the viewboxes from casting reflections on monitor screens and impairing the adaptation of physicians' eyes to the dimmer monitors.

Hardcopy Devices

Hardcopy devices permit the recording of digital images on photographic film. These devices usually provide several formats (e.g., four images per sheet, six images per sheet). Two devices for recording images on film are multiformat video cameras and laser cameras.

Multiformat Video Camera

A multiformat video camera consists of a high-quality video monitor; one or more camera lenses, each with an electrically-controlled shutter; and a holder for a cassette containing a sheet of photographic film, all in a light-tight enclosure (Fig. 17-11). The size and position of each image on the film is adjusted by moving the lens, in a single-lens system; by using different lenses, in a multiple-lens system; or by a combination of the two. A multiformat video camera is usually provided an analog video signal from a video card on an image processing workstation, and the format is manually selected by a technologist operating the camera. The spatial resolution and contrast of the video monitor limit the quality of the recorded images from a multiformat video camera.

Laser Multiformat Camera

A film laser camera consists of a laser with a modulator to vary the intensity of the beam; a rotating polygonal mirror to scan the laser beam across the film; and a film-transport mechanism to move the film in a linear fashion so that the scanning

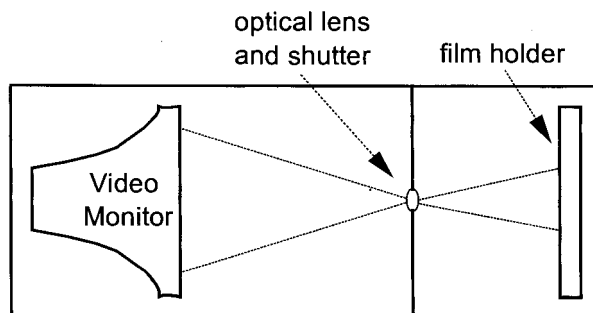


FIGURE 17-11. Multiformat video camera.

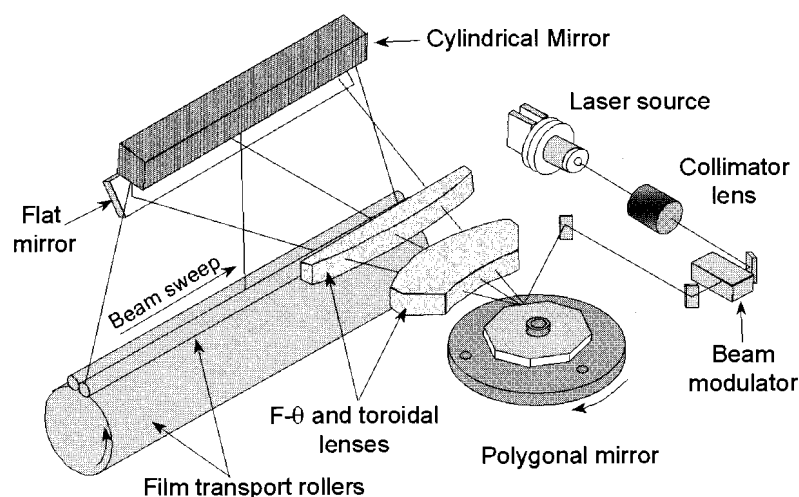


FIGURE 17-12. Schematic diagram of a film laser camera. A constant-intensity laser source is focused by a collimator lens, modulated according to the digital value of the image pixel, and directed at the film. The F-theta/toroidal lenses and the cylindrical mirror modify the shape of the beam to minimize distortion toward the periphery of the image.

process covers the entire film (Fig. 17-12). The laser camera produces images of superior resolution and dynamic range, compared with those from a video multi-format camera. The lasers in most laser cameras emit red light, requiring the use of red-sensitive film. Care must be taken not to handle this film under the normal red darkroom safelight.

Advantages and Disadvantages of PACS

There are advantages and disadvantages to PACS. Advantages include the compact storage of images on magnetic and optical storage media, rapid transfer of images for viewing at remote sites, and simultaneous access to images at different locations. Disadvantages include cost and complexity, the loss of information that occurs when analog image information is converted to digital form, and the limited dynamic range and spatial resolution of display monitors.

Advantages of PACS

1. Prompt access to images, within the medical center and especially at remote locations
2. Ability for more than one clinician to simultaneously view the same images
3. Ability to enhance images (e.g., window and level)
4. Reduced loss of clinical images
5. Reduced archival space (film file room)
6. Reduced personnel (fewer film file room clerks)
7. Reduced film costs
8. Reduced film processing
9. Computer-aided detection/diagnosis

Disadvantages of PACS

1. Initial and recurring equipment costs (e.g., replacing file cabinets with arrays of disk and tape drives, replacing lightboxes with workstations with high-performance video monitors, installing cables for networks)
2. Massive data storage requirements of radiologic images (e.g., one pair of chest images uncompressed ≈ 20 MB)
3. Expensive technical personnel to support the system
4. Loss of information when films are digitized or solid-state image receptors are substituted for film-screen combinations
5. Loss of information when lossy image compression is used
6. Lower dynamic range with video monitors than with film, when viewing an image over a similar gray-scale range, and lower spatial resolution when viewing an image without magnification
7. Physician reluctance to interpret from video monitors
8. Conversion of film to digital images
9. Provision of hardware and software to permit transfer of images between equipment from different manufacturers
10. Maintaining access to previously archived images after converting to a new archival technology
11. Security and reliability

These two lists are not all-inclusive. Of course, the desirable strategy during a PACS implementation is to enhance the advantages and minimize the disadvantages. Benefits expand as experience with these systems increase. Furthermore, the rapidly increasing capabilities and falling costs of the technology used in PACS continues to increase their performance while reducing their cost.

Security and Reliability

Two important issues regarding PACS are security and reliability. The main goals of security, as mentioned in Chapter 4 and earlier in this chapter, are to deny unauthorized persons access to confidential information (e.g., patient data) and to protect software and data from accidental or deliberate modification or loss. Methods for protection against unauthorized access were described in earlier sections. Protection against loss can be achieved by maintaining copies on more than one magnetic disk drive, tape cassette, or optical disk. The value of a study declines with time after its acquisition, and so, therefore, does the degree of protection required against its loss. A very high degree of protection is necessary until a study has been interpreted.

The goal of reliability is to ensure that acquired studies can be interpreted at all times. Reliability can be achieved by the use of reliable components and by fault-tolerant design. A design that continues to function despite equipment failure is said to be fault-tolerant. The design of a PACS must take into account the fact that equipment will fail. Fault tolerance usually implies redundancy of critical components. Not all components are critical. For example, in a PACS with a central archive connected by a network to multiple workstations, the failure of single workstation would have little adverse effect, but failure of the archive or network could prevent the interpretation of studies.

An example of a fault-tolerant strategy is to retain a copy of each study on local magnetic disk storage at the imaging device, even after a copy is sent to the central archive, until the study has been interpreted. That way, if the central archive fails, the study can still be sent over the network to an interpretation workstation, and likewise, if the network fails, a physician can view the study at the workstation of the imaging device.

Distributed PACS architectures provide a degree of fault tolerance. For example, if the nuclear medicine and ultrasound departments have separate mini-PACS interfaced to the main PACS network, failure of the main archive or network will not disrupt them.

Provisions for repair of a PACS are important. The failure of a critical component is less serious if it is quickly repaired. The time elapsed from failure of a critical component until its repair are determined by the time until the arrival of a service technician, the technician's skill in diagnosis and repair, the availability of test equipment, and the availability of parts. Appropriate arrangements for emergency service should be made in advance of equipment failure.

Quality Control of PACS and Teleradiology Systems

PACS and teleradiology systems are part of the imaging chain and therefore require acceptance testing and quality control monitoring. Although both analog and dig-

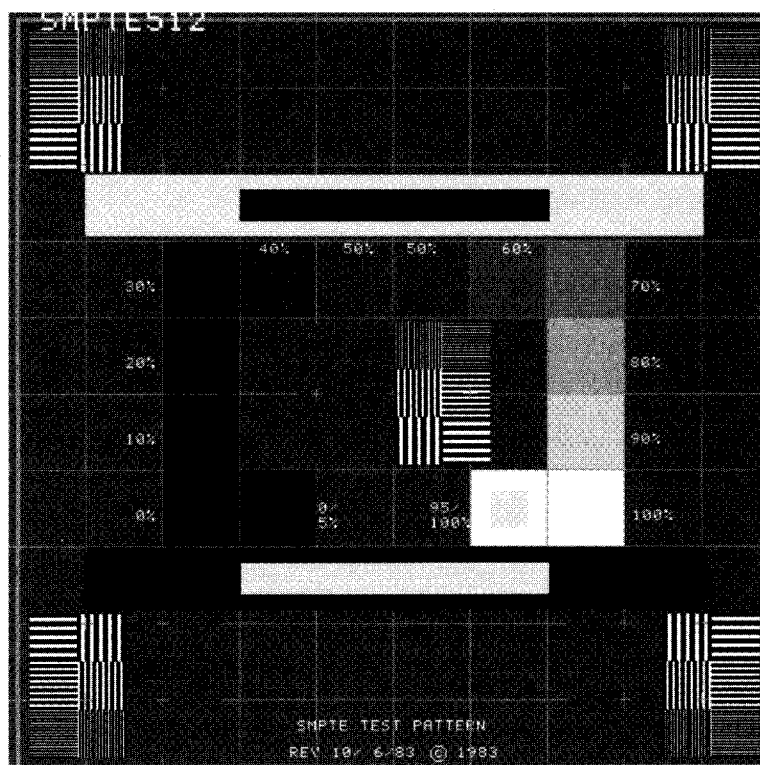


FIGURE 17-13. Society of Motion Picture Test Engineers (SMPTE) test pattern.

ital portions of the system can degrade images, the performance of digital portions (e.g., computer network, software, storage) does not usually change with time. Instead, digital portions of the system tend to operate properly or fail entirely. The performance of analog portions (film digitizers, laser printers, and video monitors) does drift and require monitoring. Quality control can be divided into acceptance testing, to determine whether newly installed components meet desired performance specifications, and routine performance monitoring. In particular, the contrast and brightness of video monitors must be evaluated periodically to ensure that all the monitors of a workstation have the same display characteristics and that they meet the requirements of the interpreting physicians. The maximal luminance of a CRT monitor degrades with the amount of time that the monitor is on and with the brightness of displayed images. Users should be trained not to leave bright images displayed any longer than necessary. Screen-saver programs, which automatically display dark patterns when users do not interact with the workstation for a period of time, can significantly prolong the lives of monitors. Maximal luminance should be measured periodically, and a monitor that becomes inadequate should be replaced.

Test patterns, such as the Society of Motion Picture and Television Engineers (SMPTE) Test Pattern, are useful for adjusting video monitors and for evaluating video monitor performance both subjectively and quantitatively (Fig. 17-13). Quantitative measurements of luminance require a calibrated photometer.

SUGGESTED READING

- American College of Radiology. ACR Standard for Teleradiology and ACR Standard for Digital Image Data Management. *Standards* 1998 Nov.
- Cusma JT, Wondrow MA, Holmes DR. Image storage considerations in the cardiac catheterization laboratory. In: *Categorical course in diagnostic radiology physics: cardiac catheterization imaging*. Radiological Society of North America, 1998.
- Honeyman JC, Frost MM, Huda W, et al. Picture archiving and communication systems (PACS). *Curr Prob Diag Radiol* 1994;4:101-160.
- Peterson LL, Davie BS. *Computer networks: a systems approach*. San Francisco: Morgan Kaufman, 1996.
- Radiological Society of North America. *A special course in computers in radiology*. Chicago: Radiological Society of North America, 1997.
- Seibert JA, Filipow LJ, Andriole KP, eds. *Practical digital imaging and PACS*. Madison, WI: Medical Physics Publishing, 1999.

SECTION
III

NUCLEAR MEDICINE

RADIOACTIVITY AND NUCLEAR TRANSFORMATION

18.1 RADIONUCLIDE DECAY TERMS AND RELATIONSHIPS

Activity

The quantity of radioactive material, expressed as the number of radioactive atoms undergoing nuclear transformation per unit time (t), is called *activity* (A). Described mathematically, activity is equal to the change (dN) in the total number of radioactive atoms (N) in a given period of time (dt), or

$$A = -dN/dt \quad [18-1]$$

The minus sign indicates that the number of radioactive atoms decreases with time. Activity is traditionally expressed in units of curies (Ci). The curie is defined as 3.70×10^{10} disintegrations per second (dps). A curie is a large amount of radioactivity. In nuclear medicine, activities from 0.1 to 30 mCi of a variety of radionuclides are typically used for imaging studies, and up to 300 mCi of iodine 131 are used for therapy. Although the curie is still the most common unit of radioactivity in the United States, the majority of the world's scientific literature uses the Systeme International (SI) units. The SI unit for radioactivity is the becquerel (Bq), named for Henri Becquerel, who discovered radioactivity in 1896. The becquerel is defined as 1 dps. One millicurie (mCi) is equal to 37 megabecquerels (1 mCi = 37 MBq).

Table 18-1 lists the units and prefixes describing various amounts of radioactivity.

Decay Constant

Radioactive decay is a random process. From moment to moment, it is not possible to predict which radioactive atoms in a sample will decay. However, observation of a larger number of radioactive atoms over a period of time allows the average rate of nuclear transformation (decay) to be established. The number of atoms decaying per unit time (dN/dt) is proportional to the number of unstable atoms (N) that are present at any given time:

$$dN/dt \propto N \quad [18-2]$$

A proportionality can be transformed into an equality by introducing a constant. This constant is called the *decay constant* (λ).

$$-dN/dt = \lambda N \quad [18-3]$$

TABLE 18-1. UNITS AND PREFIXES ASSOCIATED WITH VARIOUS QUANTITIES OF RADIOACTIVITY

Quantity	Symbol	dps (Bq)	dpm
Curie	Ci	3.7×10^{10}	2.22×10^{12}
Millicurie	mCi (10^{-3} Ci)	3.7×10^7	2.22×10^9
Microcurie	μ Ci (10^{-6} Ci)	3.7×10^4	2.22×10^6
Nanocurie	nCi (10^{-9} Ci)	3.7×10^1	2.22×10^3
Picocurie	pCi (10^{-12} Ci)	3.7×10^{-2}	2.22

The minus sign indicates that the number of radioactive atoms decaying per unit time (the decay rate or activity of the sample) decreases with time. The decay constant is equal to the fraction of the number of radioactive atoms remaining in a sample that decay per unit time. The relationship between activity and λ can be seen by considering Equation 18-1 and substituting A for $-dN/dt$ in Equation 18-3:

$$A = \lambda N \quad [18-4]$$

The decay constant is characteristic of each radionuclide. For example the decay constants for technetium-99m (Tc-99m) and molybdenum 99 (Mo-99) are 0.1151 hr^{-1} and 0.252 day^{-1} , respectively.

Physical Half-Life

A useful parameter related to the decay constant is the physical half-life ($T_{1/2}$ or $T_{p1/2}$). The half-life is defined as the time required for the number of radioactive atoms in a sample to decrease by one half. The number of radioactive atoms remaining in a sample and the number of elapsed half-lives are related by the following equation:

$$N = N_0/2^n \quad [18-5]$$

where N is number of radioactive atoms remaining, N_0 is the initial number of radioactive atoms, and n is the number of half-lives that have elapsed. The relationship between time and the number of radioactive atoms remaining in a sample is demonstrated with Tc-99m ($T_{p1/2} \approx 6$ hours) in Table 18-2.

After 10 half-lives, the number of radioactive atoms in a sample is reduced by approximately a thousand. After 20 half-lives, the number of radioactive atoms is reduced by approximately a million.

TABLE 18-2. RADIOACTIVE DECAY^a

Time (days)	No. of physical half-lives	Expression	N	$(N/N_0) \times 100 = \% \text{ remaining}$
0	0	$N_0/2^0$	10^6	100
0.25	1	$N_0/2^1$	5×10^5	50
0.5	2	$N_0/2^2$	2.5×10^5	25
0.75	3	$N_0/2^3$	1.25×10^5	12.5
1	4	$N_0/2^4$	6.25×10^4	6.25
2.5	10	$N_0/2^{10}$	$\approx 10^3$	$\approx 0.1 (10^{-3})$ or $(1/1000)N_0$
5	20	$N_0/2^{20}$	≈ 1	$\approx 0.000001 (10^{-6})$ or $(1/1,000,000)N_0$

^aThe influence of radioactive decay on the number of radioactive atoms in a sample is illustrated with technetium-99m, which has a physical half-life of 6 hours (0.25 days). The sample initially contains one million (10^6) radioactive atoms.

TABLE 18-3. PHYSICAL HALF-LIFE ($T_{p1/2}$) AND DECAY CONSTANT (λ) FOR RADIONUCLIDES USED IN NUCLEAR MEDICINE

Radionuclide	$T_{p1/2}$	λ
Fluorine 18 (^{18}F)	110 min	0.0063 m
Technetium 99m (^{99m}Tc)	6.02 hr	0.1151 hr $^{-1}$
Iodine 123 (^{123}I)	13.27 hr	0.0522 hr $^{-1}$
Samarium 153 (^{153}Sm)	1.93 d	0.3591 d $^{-1}$
Molybdenum 99 (^{99}Mo)	2.75 d	0.2522 d $^{-1}$
Indium 111 (^{111}In)	2.81 d	0.2466 d $^{-1}$
Thallium 201 (^{201}Tl)	3.04 d	0.2281 d $^{-1}$
Gallium 67 (^{67}Ga)	3.26 d	0.2126 d $^{-1}$
Xenon 133 (^{133}Xe)	5.24 d	0.1323 d $^{-1}$
Iodine 131 (^{131}I)	8.02 d	0.0864 d $^{-1}$
Phosphorus 32 (^{32}P)	14.26 d	0.0486 d $^{-1}$
Chromium 51 (^{51}Cr)	27.70 d	0.0250 d $^{-1}$
Strontium 89 (^{89}Sr)	50.53 d	0.0137 d $^{-1}$
Iodine 125 (^{125}I)	59.41 d	0.0117 d $^{-1}$
Cobalt 57 (^{57}Co)	271.79 d	0.0025 d $^{-1}$

The decay constant and the physical half-life are related as follows:

$$\lambda = \ln 2/T_{p1/2} = 0.693/T_{p1/2} \quad [18-6]$$

where $\ln 2$ denotes the natural logarithm of 2. Note that the derivation of this relationship is identical to that between the linear attenuation coefficient (μ) and the half value layer (HVL) in Chapter 3 (Equation 3-8).

The physical half-life and the decay constant are physical quantities that are inversely related and unique for each radionuclide. Half-lives of radioactive materials range from billions of years to a fraction of a second. Radionuclides used in nuclear medicine typically have half-lives on the order of hours or days. Examples of $T_{p1/2}$ and λ for radionuclides commonly used in nuclear medicine are listed in Table 18-3.

Fundamental Decay Equation

By applying the integral calculus to Equation 18-3, a useful relationship is established between the number of radioactive atoms remaining in a sample and time—the fundamental decay equation:

$$N_t = N_0 e^{-\lambda t} \text{ or } A_t = A_0 e^{-\lambda t} \quad [18-7]$$

where:

N_t = number of radioactive atoms at time t

A_t = activity at time t

N_0 = initial number of radioactive atoms

A_0 = initial activity

e = base of natural logarithm = 2.718...

λ = decay constant = $\ln 2/T_{p1/2} = 0.693/T_{p1/2}$

t = time

Problem: A nuclear medicine technologist injects a patient with 500 μCi of indium 111-labeled autologous platelets ($T_{1/2} = 2.81$ days). Two days later the patient is imaged. Assuming that none of the activity was excreted, how much activity remains at the time of imaging?

Solution:

$$A = A_0 e^{-\lambda t}$$

Given:

$$A_0 = 500 \mu\text{Ci}$$

$$\lambda = 0.693/2.81 \text{ days} = 0.246 \text{ days}^{-1}$$

$$t = 2 \text{ days}$$

Note: t and $T_{1/2}$ must be in the same units of time.

$$A_t = 500 \mu\text{Ci} e^{-(0.246 \text{ days}^{-1})(2 \text{ days})}$$

$$A_t = 500 \mu\text{Ci} e^{-0.492}$$

$$A_t = (500 \mu\text{Ci})(0.612)$$

$$A_t = 306 \mu\text{Ci}$$

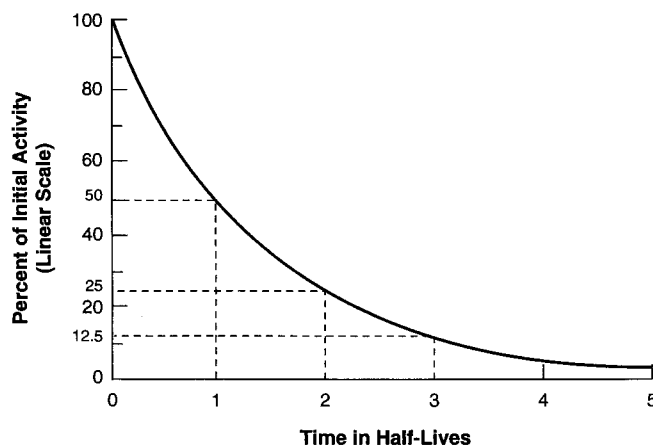


FIGURE 18-1. Percentage of initial activity as a function of time (linear plot).

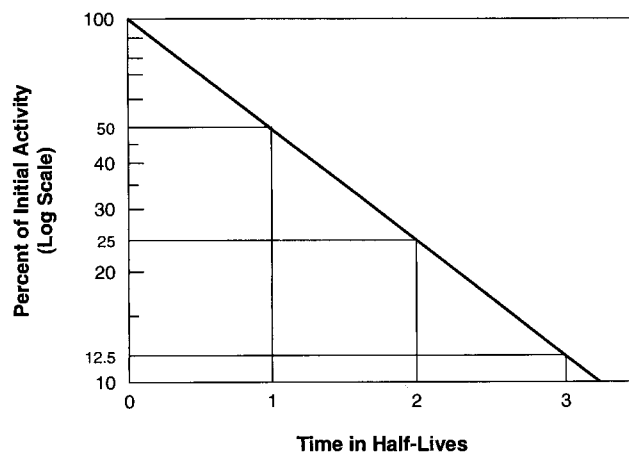


FIGURE 18-2. Percentage of initial activity as a function of time (semilog plot).

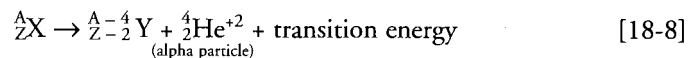
A plot on a linear axis of activity as a function of time results in a curvilinear exponential relationship in which the total activity asymptotically approaches zero (Fig. 18-1). If the logarithm of the activity is plotted versus time (semilog plot), this exponential relationship appears as a straight line (Fig. 18-2).

18.2 NUCLEAR TRANSFORMATION

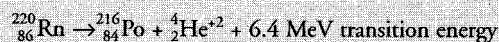
As mentioned previously, when the atomic nucleus undergoes the spontaneous transformation, called *radioactive decay*, radiation is emitted. If the daughter nucleus is stable, this spontaneous transformation ends. If the daughter is unstable (i.e., radioactive), the process continues until a stable nuclide is reached. Most radionuclides decay in one or more of the following ways: (a) alpha decay, (b) beta-minus emission, (c) beta-plus (positron) emission, (d) electron capture, or (e) isomeric transition.

Alpha Decay

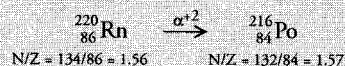
Alpha (α) decay is the spontaneous emission of an alpha particle (identical to a helium nucleus consisting of two protons and two neutrons) from the nucleus. Alpha decay typically occurs with heavy nuclides ($A > 150$) and is often followed by gamma and characteristic x-ray emission. These photon emissions are often accompanied by the competing processes of internal conversion and Auger electron emission. Alpha particles are the heaviest and least penetrating form of radiation consider in this chapter. They are emitted from the atomic nucleus with discrete energies in the range of 2 to 10 MeV. An alpha particle is approximately four times heavier than a proton or neutron and carries an electronic charge twice that of the proton. Alpha decay can be described by the following equation:



Example:



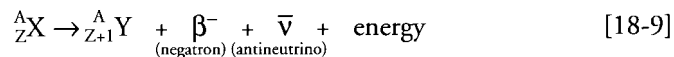
Alpha decay results in a large energy transition and a slight increase in the ratio of neutrons to protons (N/Z ratio):



Alpha particles are not used in medical imaging because their ranges are limited to approximately 1 cm/MeV in air and typically less than 100 μm in tissue. Even the most energetic alpha particles cannot penetrate the dead layer of the skin. However, the intense ionization tracks produced by alpha particles make them a serious health hazard should alpha-emitting radionuclides enter the body via ingestion, inhalation, or a wound. Research efforts are underway to assess the clinical utility of alpha-emitting radionuclides chelated to monoclonal antibodies directed against various tumors as radioimmunotherapeutic agents.

Beta-Minus (Negatron) Decay

Beta-minus (β^-) decay, or negatron decay, characteristically occurs with radionuclides that have an excess number of neutrons compared with the number of protons (i.e., a high N/Z ratio). Beta-minus decay can be described by the following equation:



This mode of decay results in the conversion of a neutron into a proton with the simultaneous ejection of a negatively charged beta particle (β^-) and an antineutrino ($\bar{\nu}$). With the exception of their origin (the nucleus), beta particles are identical to ordinary electrons. The antineutrino is an electrically neutral subatomic particle whose mass is much smaller than that of an electron. The absence of charge and the infinitesimal mass of antineutrinos make them very difficult to detect because they rarely interact with matter. Beta decay increases the number of protons by 1 and thus transforms the atom into a different element with an atomic number $Z + 1$. However, the concomitant decrease in the neutron number means that the mass number remains unchanged. Decay modes in which the mass number remains constant are called *isobaric transitions*. Radionuclides produced by nuclear fission are “neutron rich,” and therefore most decay by β^- emission. Beta-minus decay decreases the N/Z ratio, bringing the daughter closer to the line of stability (see Chapter 2):

Example:

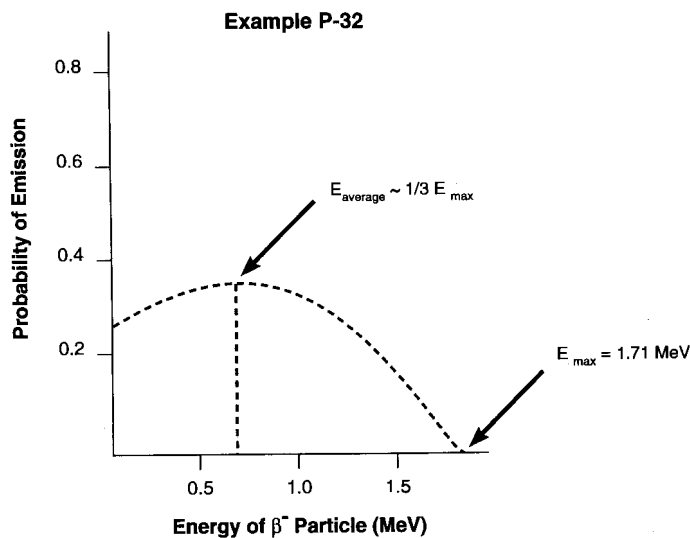
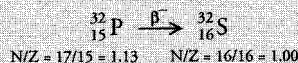
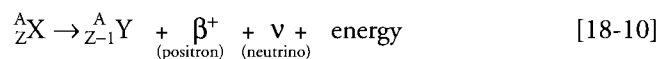


FIGURE 18-3. Distributions of energies of beta particles.

Although the β^- particles emitted by a particular radionuclide have a discrete maximal energy ($E_{\max}$), most are emitted with energies lower than the maximum. The average energy of the β^- particles is approximately $1/3 E_{\max}$. The balance of the energy is given to the antineutrino (i.e., $E_{\max} = E_{\beta^-} + E_{\bar{\nu}}$). Thus, beta-minus decay results in a polyenergetic spectrum of β^- energies ranging from zero to $E_{\max}$ (Fig. 18-3). Any excess energy in the nucleus after beta decay is emitted as gamma rays, internal conversion electrons, and other associated radiations.

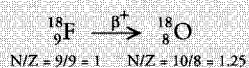
Beta-Plus Decay (Positron Emission)

Just as beta-minus decay is driven by the nuclear instability caused by excess neutrons, “neutron-poor” radionuclides (i.e., those with a low N/Z ratio) are also unstable. Many of these radionuclides decay by beta-plus (positron) emission, which increases the neutron number by one. Beta-plus decay can be described by the following equation:



The net result is the conversion of a proton into a neutron with the simultaneous ejection of the positron (β^+) and a neutrino (ν). Positron decay decreases the number of protons (atomic number) by 1 and thereby transforms the atom into a different element with an atomic number of $Z-1$. The number of neutrons is increased by 1; therefore, the transformation is isobaric because the total number of nucleons is unchanged. Accelerator-produced radionuclides, which are typically neutron deficient, often decay by positron emission. Positron decay increases the N/Z ratio, resulting in a daughter closer to the line of stability.

Example:



The energy distribution between the positron and the neutrino is similar to that between the negatron and the antineutrino in beta-minus decay; thus positrons are polyenergetic with an average energy equal to approximately $1/3 E_{\max}$. As with β^- decay, excess energy following positron decay is released as gamma rays and other associated radiation.

Although β^+ decay has similarities to β^- decay, there are also important differences. The neutrino and antineutrino are *antiparticles*, as are the positron and negatron. The prefix *anti-* before the name of an elementary particle denotes another particle with certain symmetry characteristics. In the case of charged particles such as the positron, the antiparticle (i.e., the negatron) has a charge equal but opposite to that of the positron and a magnetic moment that is oppositely directed with respect to spin. In the case of neutral particles such as the neutrino and antineutrino, there is no charge; therefore, differentiation between the particles is made solely on the basis of differences in magnetic moment. Other important differences between the particle and antiparticle are their lifetimes and their eventual fates. As mentioned earlier, negatrons are physically identical to ordinary electrons and as such lose their kinetic energy as they traverse matter via excitation and ionization.

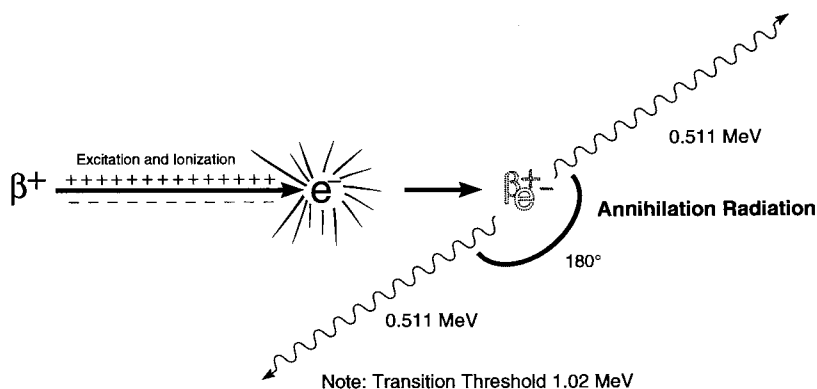
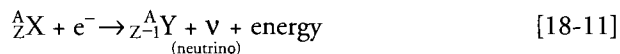


FIGURE 18-4. Annihilation radiation.

When they lose all (or most) of their kinetic energy, they may be captured by an atom or absorbed into the free electron pool. Positrons undergo a similar process of energy deposition via excitation and ionization; however, when they come to rest they react violently with their antiparticles (electrons). This process results in the entire rest mass of both particles being instantaneously converted to energy and emitted as two oppositely directed (i.e., 180 degrees apart) 511-keV annihilation photons (Fig. 18-4). According to Einstein's mass-energy equivalence formula, $E = mc^2$, 511 keV is the energy equivalent of the rest mass of an electron (positron or negatron). Therefore, there is an inherent threshold for positron decay equal to the sum of the annihilation photon energies (i.e., 2×511 keV, or 1.02 MeV). The transition energy between the parent and daughter nuclide must be greater than or equal to 1.02 MeV for positron decay to occur. Medical imaging of annihilation radiation from positron-emitting radiopharmaceuticals, called positron emission tomography (PET), is discussed in Chapter 22.

Electron Capture Decay

Electron capture (ϵ) is an alternative to positron decay for neutron-deficient radionuclides. In this decay mode, the nucleus captures an orbital (usually a *K*- or *L*-shell) electron, with the conversion of a proton into a neutron and the simultaneous ejection of a neutrino. Electron capture can be described by the following equation:



The net effect of electron capture is the same as positron emission: the atomic number is decreased by 1, creating a different element, and the mass number remains unchanged. Therefore, electron capture is isobaric and results in an increase in the N/Z ratio.

Example:



The capture of an orbital electron creates a vacancy in the electron shell, which is filled by an electron from a higher-energy shell. As discussed in Chapter 2, this electron transition results in the emission of characteristic x-rays and/or Auger electrons. For example, thallium 201 (Tl-201) decays to mercury-201 (Hg-201) by electron capture, resulting in the emission of characteristic x-rays. It is these x-rays that are primarily used to create the images in Tl-201 myocardial perfusion studies. As with other modes of decay, if the nucleus is left in an excited state following electron capture, the excess energy will be emitted as gamma rays and other radiations.

As previously mentioned, positron emission requires an energy difference between the parent and daughter atoms of at least 1.02 MeV. Neutron-poor radionuclides below this threshold transition energy decay exclusively by electron capture. Nuclides with parent-to-daughter transition energies greater than 1.02 MeV may decay by electron capture or positron emission, or both. Heavier proton-rich nuclides are more likely to decay by electron capture, whereas lighter proton-rich nuclides are more likely to decay by positron emission. This is a result of the closer proximity of the *K*- or *L*-shell electrons to the nucleus and the greater magnitude of the coulombic attraction from the positive charges. Although capture of a *K*- or *L*-shell electron is the most probable, electron capture can occur with higher-energy shell electrons.

The quantum-mechanical description of the atom is essential for understanding electron capture. The Bohr model describes electrons in fixed orbits at discrete distances from the nucleus. This model does not permit electrons to be close enough to the nucleus to be captured. However, the quantum-mechanical model describes orbital electron locations as probability density functions in which there is a finite probability that an electron will pass close to or even through the nucleus.

Electron capture radionuclides used in medical imaging decay to atoms in excited states that subsequently emit externally detectable x-rays or gamma rays or both.

Isomeric Transition

Often, during radioactive decay (i.e., α^{++} , β^- , β^+ , or ϵ emission), a daughter is formed in an excited (i.e., unstable) state. Gamma rays are emitted as the daughter nucleus undergoes an internal rearrangement and transitions from the excited state to a lower-energy state.

Once created, most excited states transition almost instantaneously to lower-energy states with the emission of gamma radiation. However, some excited states persist for longer periods, with half-lives ranging from approximately 10^{-12} seconds to more than 600 years. These excited states are called metastable or isomeric states and are denoted by the letter “m” after the mass number (e.g., Tc-99m). Isomeric transition is a decay process that yields gamma radiation without the emission or capture of a particle by the nucleus. There is no change in atomic number, mass number, or neutron number. Thus, this decay mode is isobaric and isotonic, and it occurs between two nuclear energy states with no change in the *N/Z* ratio.

Isomeric transition can be described by the following equation:



The energy is released in the form of gamma rays or internal conversion electrons, or both.

Decay Schemes

Each radionuclide's decay process is a unique characteristic of that radionuclide. The majority of the pertinent information about the decay process and its associated radiation can be summarized in a line diagram called a *decay scheme* (Fig. 18-5). Decay schemes identify the parent, daughter, mode of decay, intermediate excited states, energy levels, radiation emissions, and sometimes physical half-life. The top horizontal line represents the parent, and the bottom horizontal line represents the daughter. Horizontal lines between those two represent intermediate excited states. A diagonal line to the left is used to indicate electron capture decay; a short vertical line followed by a diagonal line to the left indicates either positron or alpha decay; and a diagonal line to the right indicates beta-minus decay. Vertical lines indicate gamma ray emission, including isomeric transition. These diagrams are often accompanied by decay data tables, which provide information on all the significant ionizing radiations emitted from the atom as a result of the nuclear transformation. Examples of these decay schemes and data tables are presented in this section.

Figure 18-6 shows the alpha decay scheme of radon-220 (Rn-220). Rn-220 has a physical half-life of 55 seconds and decays by one of two possible alpha transitions. Alpha 1 (α_1) at 5.747 MeV occurs 0.07% of the time and is followed immediately by a 0.55-MeV gamma ray (γ_1) to the ground state. The emission of alpha 2 (α_2), with an energy of 6.287 MeV, occurs 99.3% of the time and leads directly to the ground state. The decay data table lists these radiations together with the daughter atom, which has a -2 charge and a small amount of kinetic energy as a result of alpha particle emission.

Phosphorus 32 (P-32) is used in nuclear medicine as a therapeutic agent in the treatment of a variety of diseases, including polycythemia vera, metastatic bone disease, and serous effusions. P-32 has a half-life of 14.3 days and decays directly to its ground state by emitting a beta-minus particle with an $E_{\max}$ of 1.71 MeV (Fig. 18-7). The average (mean) energy of the beta-minus particle is approximately $1/3 E_{\max}$ (0.6948 MeV), with the antineutrino carrying off the balance of the transition energy. There are no excited energy states or other radiation emitted during this decay; therefore, P-32 is referred to as a "pure beta emitter."

A somewhat more complicated decay scheme is associated with the beta-minus decay of Mo-99 to Tc-99 (Fig. 18-8). In this case there are eight possible beta-minus decay transitions with probabilities ranging from 0.797 for beta-minus 8 (i.e., 79.7% of all decays of Mo-99 are by β_8^- transition) to 0.0004 (0.04%) for beta-minus 6. The sum of all transition probabilities (β_1^- to β_8^-) is equal to 1. The aver-

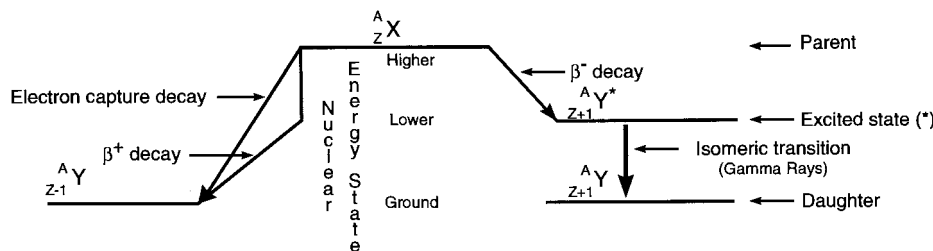
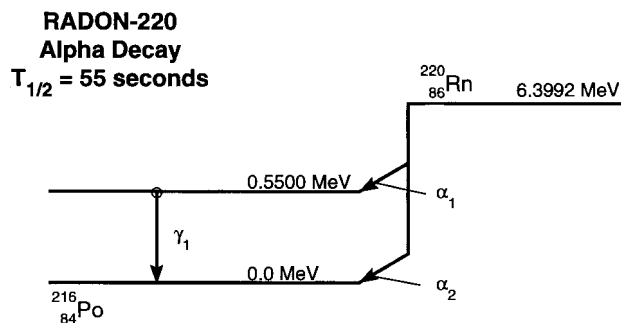


FIGURE 18-5. Elements of the generalized decay scheme.



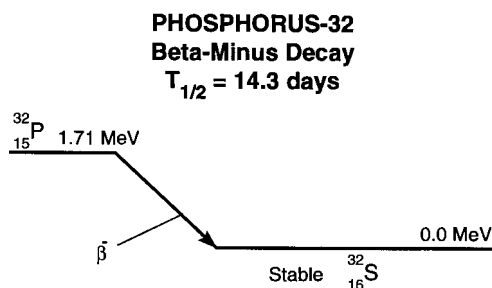
Decay Data Table

Radiation		Mean Number per Disintegration	Mean Energy per Particle (MeV)
Alpha	1	0.0007	5.7470
Recoil Atom		0.0007	0.1064
Alpha	2	0.9993	6.2870
Recoil Atom		0.9993	0.1164
Gamma	1	0.0006	0.5500

FIGURE 18-6. Principal decay scheme of radon 220.

age energy of beta particles from the transition is 0.4519 MeV. The transition leads to a metastable state of technetium 99, Tc-99m, which is 0.1426 MeV above the ground state and decays with a half-life of 6.02 hours. Tc-99m is the most widely used radionuclide in nuclear medicine.

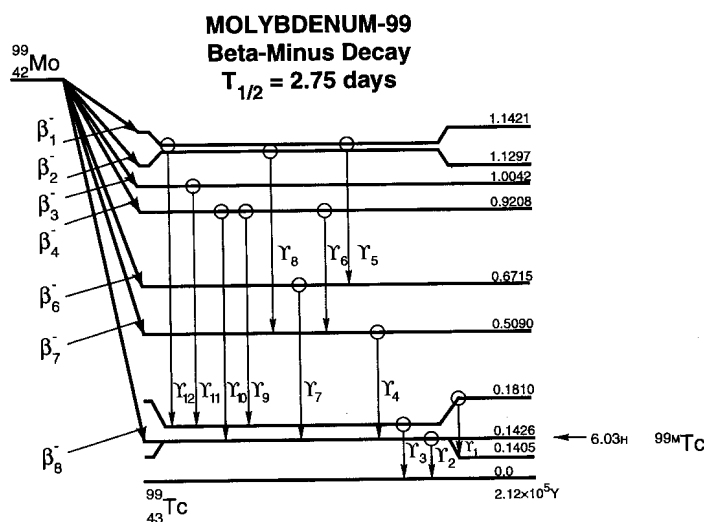
After beta decay, there are a number of excited states created that transition to lower-energy levels via the emission of gamma rays and/or internal conversion elec-



Decay Data Table

Radiation	Mean Number per Disintegration	Mean Energy per Particle (MeV)
Beta Minus	1.000	0.6948

FIGURE 18-7. Principal decay scheme of phosphorus 32.



Decay Data Table

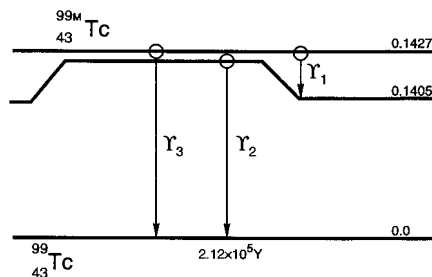
Mean			Mean Energy	Mean			Mean Energy
		Number per	per Particle		Number per	per Particle	
Radiation		Disintegration	(MeV)	Radiation	Disintegration	(MeV)	
Beta Minus	1	0.0012	0.0658	Gamma	4	0.0143	0.3664
Beta Minus	3	0.0014	0.1112	Gamma	5	0.0001	0.3807
Beta Minus	4	0.1850	0.1401	Gamma	6	0.0002	0.4115
Beta Minus	6	0.0004	0.2541	Gamma	7	0.0005	0.5289
Beta Minus	7	0.0143	0.2981	Gamma	8	0.0002	0.6207
Beta Minus	8	0.7970	0.4519	Gamma	9	0.1367	0.7397
Gamma	1	0.0130	0.0405	K Int Con Elect		0.0002	0.7186
K Int Con Elect		0.0428	0.0195	Gamma	10	0.0479	0.7782
L Int Con Elect		0.0053	0.0377	K Int Con Elect		0.0000	0.7571
M Int Con Elect		0.0017	0.0401	Gamma	11	0.0014	0.8231
Gamma	2	0.0564	0.1405	Gamma	12	0.0011	0.9610
K Int Con Elect		0.0058	0.1194	K Alpha-1 X-Ray		0.0253	0.0183
L Int Con Elect		0.0007	0.1377	K Alpha-2 X-Ray		0.0127	0.0182
Gamma	3	0.0657	0.1810	K Beta-1 X-Ray		0.0060	0.0206
K Int Con Elect		0.0085	0.1600	KLL Auger Elect		0.0087	0.0154
L Int Con Elect		0.0012	0.1782	KLX Auger Elect		0.0032	0.0178
M Int Con Elect		0.0004	0.1806	LMM Auger Elect		0.0615	0.0019
				MXY Auger Elect		0.1403	0.0004

FIGURE 18-8. Principal decay scheme of molybdenum-99.

trons. As previously described, the ejection of an electron by internal conversion of the gamma ray results in the emission of characteristic x-rays or Auger electrons, or both. All of these radiations, their mean energies, and their associated probabilities are included in the decay data table.

The process of gamma ray emission by isomeric transition is of primary importance to nuclear medicine, because most procedures performed depend on the emission and detection of gamma radiation. Figure 18-9 shows the decay scheme for Tc-99m. There are three gamma-ray transitions as Tc-99m decays to Tc-99. The gamma 1 (γ_1) transition occurs 98.6% of the time; however, no 2.2-keV gamma rays are emitted because virtually all of this energy is internally converted resulting in the emission of *M*-shell internal conversion electrons with a mean energy of 1.6

TECHNETIUM ^{99m}Tc
Isomeric Transition
 $T_{1/2} = 6.02 \text{ hrs.}$



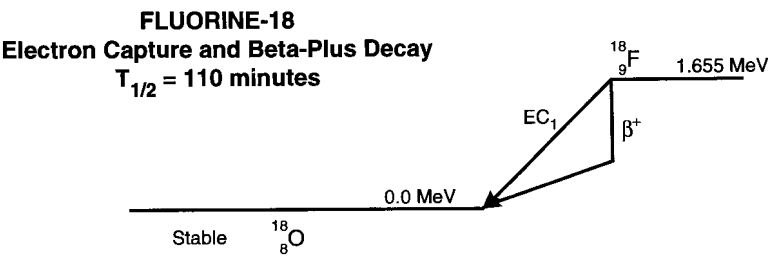
Decay Data Table

Radiation		Mean Number per Disintegration	Mean Energy per Particle (MeV)
Gamma	1	0.0000	0.0021
M Int Con Elect		0.9860	0.0016
Gamma	2	0.8787	0.1405
K Int Con Elect		0.0913	0.1194
L Int Con Elect		0.0118	0.1377
M Int Con Elect		0.0039	0.1400
Gamma	3	0.0003	0.1426
K Int Con Elect		0.0088	0.1215
L Int Con Elect		0.0035	0.1398
M Int Con Elect		0.0011	0.1422
K Alpha-1 X-Ray		0.0441	0.0183
K Alpha-2 X-Ray		0.0221	0.0182
K Beta-1 X-Ray		0.0105	0.0206
KLL Auger Elect		0.0152	0.0154
KLX Auger Elect		0.0055	0.0178
LMM Auger Elect		0.1093	0.0019
MXY Auger Elect		1.2359	0.0004

FIGURE 18-9. Principal decay scheme of technetium-99m.

keV. After internal conversion, the nucleus is left in an excited state, which is followed almost instantaneously by gamma 2 (γ_2) transition at 140.5 keV to ground state. Although the γ_2 transition occurs 98.6% of the time, the decay data table shows that the gamma ray is emitted only 87.87% of the time, with the balance of the transitions occurring via internal conversion. Gamma 2 is the principal photon imaged in nuclear medicine. The gamma 3 (γ_3) transition at 142.7 keV occurs only 1.4% of the time and is followed, almost always, by internal conversion electron emission. Conversion electrons, characteristic x-rays, and Auger electrons are also emitted as the energy from some of the transitions are internally converted.

As discussed previously, positron emission and electron capture are competing decay processes for neutron-deficient radionuclides. As shown in Fig. 18-10, fluorine-18 (F-18) decays by both modes. F-18 decays by positron emission 97% of the time and by electron capture 3% of the time. The dual mode of decay results in an "effective" decay constant (λ_e) that is the sum of the positron (λ_1) and electron capture (λ_2) decay constants: $\lambda_e = \lambda_1 + \lambda_2$. The decay data table shows that positrons are emitted 97% of the time with an average energy of 0.2496 MeV. Furthermore, the interaction of the positron with an electron results in the production of two 511 keV annihilation radiation photons. Because two photons are produced for each positron, their abundance is $2 \times 97\%$, or 194%. ^{18}F is the most widely used radionuclide for PET imaging.



Decay Data Table		
Radiation	Mean Number per Disintegration	Mean Energy Particle (MeV)
Beta Plus	0.9700	0.2496
Annih. Radiation	1.9400	0.5110

FIGURE 18-10. Principal decay scheme of fluorine 18.

SUGGESTED READING

Friedlander G, Kennedy JW, Miller JM. *Nuclear and radiochemistry*, 3rd ed. New York: Wiley, 1981.

Patton JA. Introduction to nuclear physics. *Radiographics* 1998;18:995–1007.

Sorensen JA, Phelps ME. *Physics in nuclear medicine*, 2nd ed. New York: Grune & Stratton, 1987.

RADIONUCLIDE PRODUCTION AND RADIOPHARMACEUTICALS

19.1 RADIONUCLIDE PRODUCTION

Although many naturally occurring radioactive nuclides exist, all of those commonly administered to patients in nuclear medicine are artificially produced. Artificial radioactivity was discovered by Irene Curie (daughter of Marie Curie) and Frederic Joliot in 1934, who induced radioactivity in boron and aluminum targets by bombarding them with alpha (α) particles from polonium. Positrons continued to be emitted from the targets after the alpha source was removed. Today, more than 2,500 artificial radionuclides have been produced by a variety of methods. Most radionuclides used in nuclear medicine are produced by cyclotrons, nuclear reactors, or radionuclide generators.

Cyclotron-Produced Radionuclides

Cyclotrons and other charged-particle accelerators produce radionuclides by bombarding stable nuclei with high-energy charged particles. Protons, deuterons (^2H nuclei), tritons (^3H nuclei), and alpha particles (^4He nuclei) are commonly used to produce radionuclides used in medicine. Heavy charged particles must be accelerated to high kinetic energies to overcome and penetrate the repulsive coulombic barrier of the target atoms' nuclei. In 1930, Cockcroft and Walton applied a clever scheme of cascading a series of transformers, each capable of stepping up the voltage by several hundred thousand volts. The large potentials generated were used to produce artificial radioactivity by accelerating particles to high energies and bombarding stable nuclei. In Berkeley, California, E.O. Lawrence capitalized on this development in 1931 but added a unique dimension in his design of the cyclotron (Fig. 19-1). Positive ions injected into the center of the cyclotron are attracted to and accelerated toward a negatively charged, hollow, semicircular electrode shaped like and called a "dee." There are two dees separated by a small gap and kept under an extremely high vacuum. Constrained by a static magnetic field, the ions travel in a circular path, where the radius of the circle increases as the ions are accelerated (Fig. 19-2). Half way around the circle, the ions approach the other dee and, at the same instant, the polarity of the electrical field between the two dees is reversed, causing the ions to continue their acceleration toward the negative potential. This alternating potential between the dees is repeated again and again as the particles

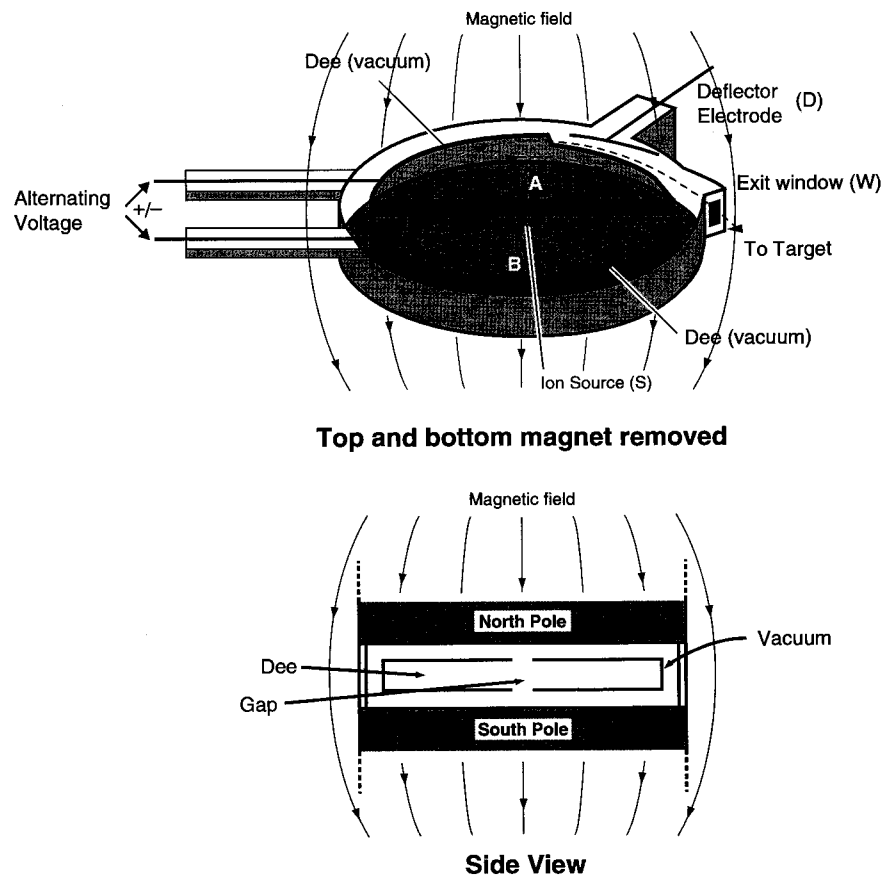


FIGURE 19-1. Schematic view of a cyclotron.

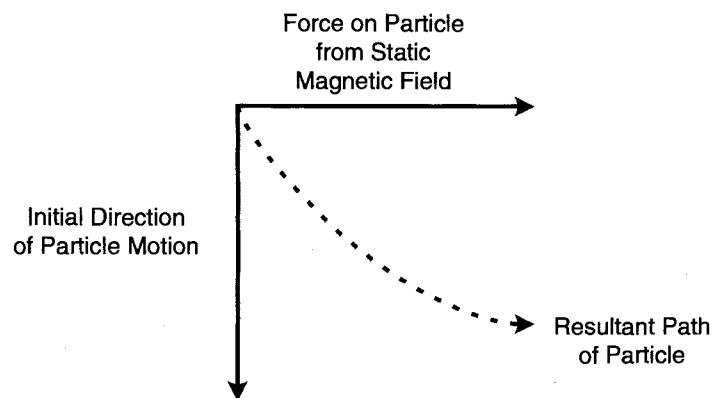


FIGURE 19-2. Forces from the static magnetic field cause the ions to travel in a circular path.

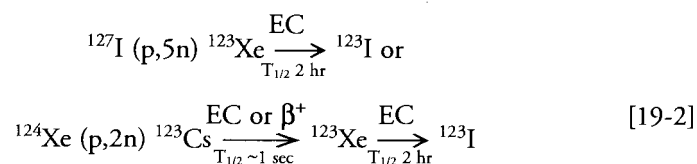
continue to acquire kinetic energy, sweeping out larger and larger circles. As the length of the pathway between successive accelerations increases, the speed of the particle also increases; hence, the time interval between accelerations remains constant. The cyclic nature of these events led to the name "cyclotron." The final energy achieved by the accelerated particles depends on the diameter of the dees and on the strength of the static magnetic field. Finally, as the ions reach the periphery of the dees, they are removed from their circular path by a negatively charged deflector plate, emerge through the window, and strike the target. Depending on the design of the cyclotron, particle energies can range from a few million electron volts (MeV) to several billion electron volts (BeV).

The accelerated ions collide with the target nuclei, causing nuclear reactions. An incident particle may leave the target nucleus after interacting, transferring some of its energy to it, or it may be completely absorbed. The specific reaction depends on the type and energy of the bombarding particle as well as the composition of the target. In any case, target nuclei are left in excited states, and this excitation energy is disposed of by the emission of particulate (protons and neutrons) and electromagnetic (gamma rays) radiations. Gallium-67 (Ga-67) is an example of a widely used cyclotron-produced radionuclide. The production reaction is written as follows:

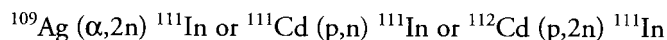


where the target material is zinc-68 (Zn-68), the bombarding particle is a proton (p) accelerated to approximately 20 MeV, two neutrons (2n) are emitted, and Ga-67 is the product radionuclide. In some cases, the nuclear reaction produces a radionuclide that decays to the clinically useful radionuclide (see iodine-123 and thallium-201 production below). Most cyclotron-produced radionuclides are neutron poor and therefore decay by positron emission or electron capture. The production methods of several cyclotron-produced radionuclides important to nuclear medicine are listed, (EC = electron capture, $T_{1/2}$ = physical half-life).

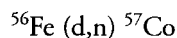
Iodine-123 production:



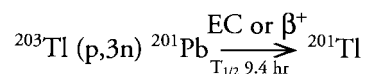
Indium-111 production:



Cobalt-57 production:



Thallium-201 production:



Industrial cyclotron facilities that produce large activities of clinically useful radionuclides are very expensive and require substantial cyclotron and radiochem-

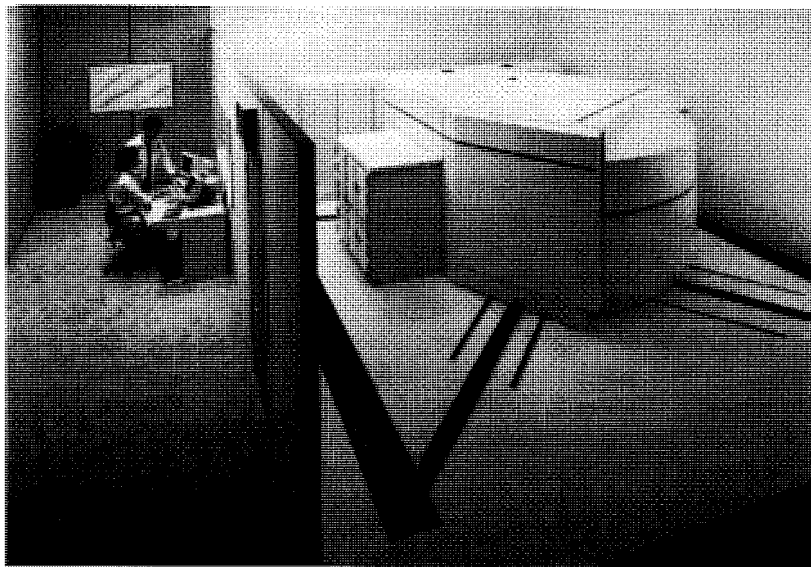


FIGURE 19-3. Hospital-based cyclotron facility. (Courtesy of Nuclear Medicine Division, Siemens Medical Systems, Hoffmann Estates, Illinois.)

istry support staff and facilities. Cyclotron-produced radionuclides are usually more expensive than those produced by other technologies.

Much smaller, specialized hospital-based cyclotrons have been developed to produce positron-emitting radionuclides for positron emission tomography (PET) (Fig. 19-3). Production methods of clinically useful positron-emitting radionuclides are listed below.

Fluorine-18 production: ^{18}O (p,n) ^{18}F ($T_{1/2} = 110$ min)

Nitrogen-13 production: ^{12}C (d,n) ^{13}N ($T_{1/2} = 10$ min)

Oxygen-15 production: ^{14}N (d,n) ^{15}O or ^{15}N (p,n) ^{15}O ($T_{1/2} = 2.0$ min) [19-3]

Carbon-11 production: ^{10}B (d,n) ^{11}C ($T_{1/2} = 20.4$ min)

The medical cyclotrons are usually located near the PET imager because of the short half-lives of the radionuclides produced. Fluorine-18 (F-18) is an exception to this generalization because of its longer half-life (110 minutes).

Nuclear Reactor–Produced Radionuclides

Nuclear reactors are another major source of clinically used radionuclides. Neutrons, being uncharged, have an advantage in that they can penetrate the nucleus without being accelerated to high energies. There are two principal methods by which radionuclides are produced in a reactor: nuclear fission and neutron activation.

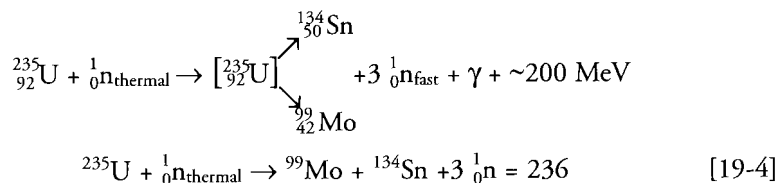
Nuclear Fission

Fission is the splitting of an atomic nucleus into two smaller nuclei. Whereas some unstable nuclei fission spontaneously, others require the input of energy to over-

come the nuclear binding forces. This energy is often provided by the absorption of neutrons. Neutrons can induce fission only in certain very heavy nuclei. Whereas high-energy neutrons can induce fission in several such nuclei, there are only three nuclei of reasonably long half-life that are fissionable by neutrons of all energies; these are called fissile nuclides.

The most widely used fissile nuclide is uranium 235 (U-235). Elemental uranium exists in nature primarily as U-238 (99.3%) with a small fraction of U-235 (0.7%). U-235 has a high fission cross-section (i.e., high fission probability); therefore, its concentration in uranium ore is usually enriched (typically to 3% to 5%) to make the fuel used in nuclear reactors.

When a U-235 nucleus absorbs a neutron, the resulting nucleus (U-236) is in an extremely unstable excited energy state that usually promptly fissions into two smaller nuclei called *fission fragments*. The fission fragments fly apart with very high kinetic energies, with the simultaneous emission of gamma radiation and the ejection of two to five neutrons per fission (Equation 19-4).



The fission of uranium creates fission fragment nuclei having a wide range of mass numbers. More than 200 radionuclides with mass numbers between 70 and 160 are produced by the fission process (Fig. 19-4). These fission products are neutron rich and therefore almost all of them decay by beta-minus (β^-) particle emission.

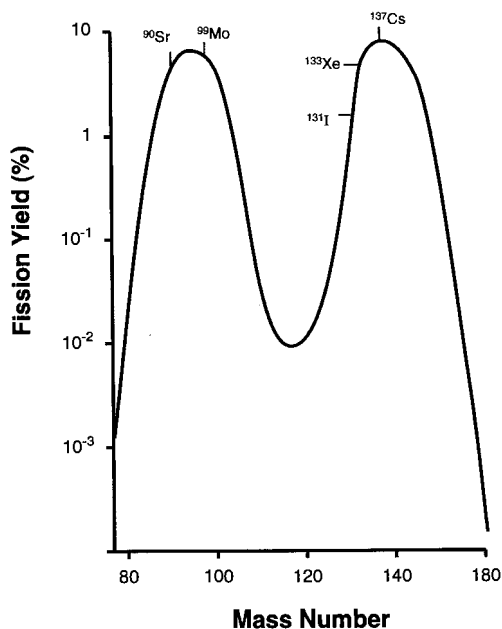


FIGURE 19-4. Fission yield as a percentage of total fission products from uranium 235.

Nuclear Reactors and Chain Reactions

The total energy released by the nuclear fission of a uranium atom is more than 200 MeV. Under the right conditions, this reaction can be perpetuated if the fission neutrons interact with other U-235 atoms, causing additional fissions and leading to a self-sustaining nuclear chain reaction (Fig. 19-5). The probability of fission (called the *fission cross-section*) with U-235 is greatly enhanced as neutrons slow down or *thermalize*. The neutrons emitted from the fission are very energetic (called *fast neutrons*), and are slowed (*moderated*) to thermal energies (~ 0.025 eV) as they scatter in water in the reactor core. Good moderators are low-Z materials that slow the neutrons without absorbing a significant fraction of them. Water is the most commonly used moderator, although other materials, such as graphite (used in the reactors at the Chernobyl plant in the Ukraine) and heavy water ($^2\text{H}_2\text{O}$), are also used.

Some neutrons are absorbed by nonfissionable material in the reactor, while others are moderated and absorbed by U-235 atoms and induce additional fissions. The ratio of the number of fissions in one generation to the number in the previous generation is called the *multiplication factor*. When the number of fissions per generation is constant, the multiplication factor is 1 and the reactor is said to be *critical*. When the multiplication factor is greater than 1, the rate of the chain reaction increases, at which time the reactor is said to be *supercritical*. If the multiplication factor is less than 1 (i.e., more neutrons being absorbed than pro-

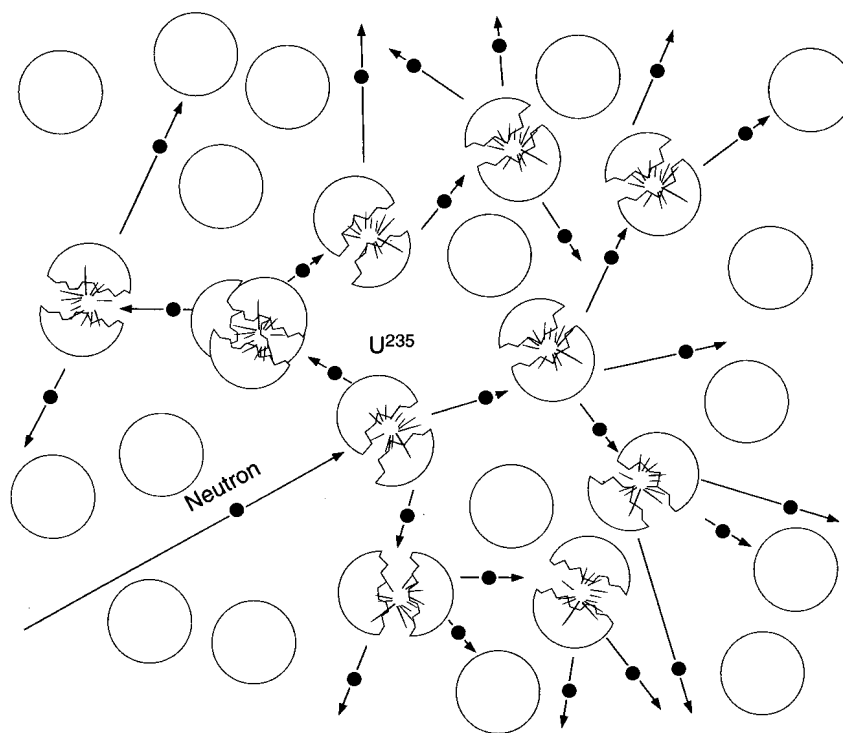


FIGURE 19-5. Schematic of a nuclear chain reaction. The neutrons (shown as small blackened circles) are not drawn to scale with respect to the uranium atoms.

duced), the reactor is said to be *subcritical* and the chain reaction will eventually cease.

This chain reaction process is analogous to a room whose floor is filled with mousetraps, each one having a ping pong ball placed on the trap. Without any form of control, a self-sustaining chain reaction will be initiated when a single ping pong ball is tossed into the room and springs one of the traps. The nuclear chain reaction is maintained at a desired level by limiting the number of available neutrons through the use of neutron-absorbing *control rods* (containing boron, cadmium, indium, or a mixture of these elements), which are placed in the reactor core between the fuel elements. Inserting the control rods deeper into the core absorbs more neutrons, reducing the reactivity (i.e., causing the neutron flux and power output to decrease with time). Removing the control rods has the opposite effect. If a nuclear reactor accident results in rapid loss of the coolant, the fuel can overheat and melt (so-called *meltdown*). However, because of the design characteristics of the reactor and its fuel, an atomic explosion, like those from nuclear weapons, is impossible.

Figure 19-6 is a diagram of a typical research/radionuclide production reactor. The fuel is processed into rods of uranium-aluminum alloy approximately 6 cm in diameter and 2.5 m long. These *fuel rods* are encased in zirconium or aluminum, which have favorable neutron and heat transport properties. There may be as many as 1,000 fuel rods in the reactor, depending on the design and the neutron flux requirements. Water circulates between the encased fuel rods in a closed loop where the heat generated from the fission process is transferred to cooler water in the heat exchanger. The water in the reactor and in the heat exchanger are closed loops that do not come into direct physical contact with the fuel. The heat transferred to the cooling water is released to the environment through cooling towers or evaporation ponds. The cooled water is pumped back toward the fuel rods, where it is reheated and the process is repeated.

In commercial nuclear power electric generation stations, the heat generated from the fission process produces pressurized steam that is directed through a steam turbine, which powers an electrical generator. The steam is then condensed to water by the condenser.

Nuclear reactor safety design principles dictate numerous barriers between the radioactivity in the core and the environment. For example, in commercial power reactors the fuel is encased in metal fuel rods that are surrounded by water and enclosed in a sealed, pressurized, 30-cm-thick steel reactor vessel. These components, together with other highly radioactive reactor systems, are enclosed in a large steel-reinforced concrete shell (approximately 1 to 2 m thick), called the *containment structure*. In addition to serving as a moderator and coolant, the water in the reactor acts as a radiation shield, reducing the radiation levels adjacent to the reactor vessel.

Fission-Produced Radionuclides

Specialized nuclear reactors are used to produce clinically useful radionuclides from fission products or neutron activation of stable target material. Ports exist in the reactor core between the fuel elements where samples to be irradiated are inserted. The fission products most often used in nuclear medicine are molybdenum-99 (Mo-99), iodine-131 (I-131), and xenon-133 (Xe-133). These products can be

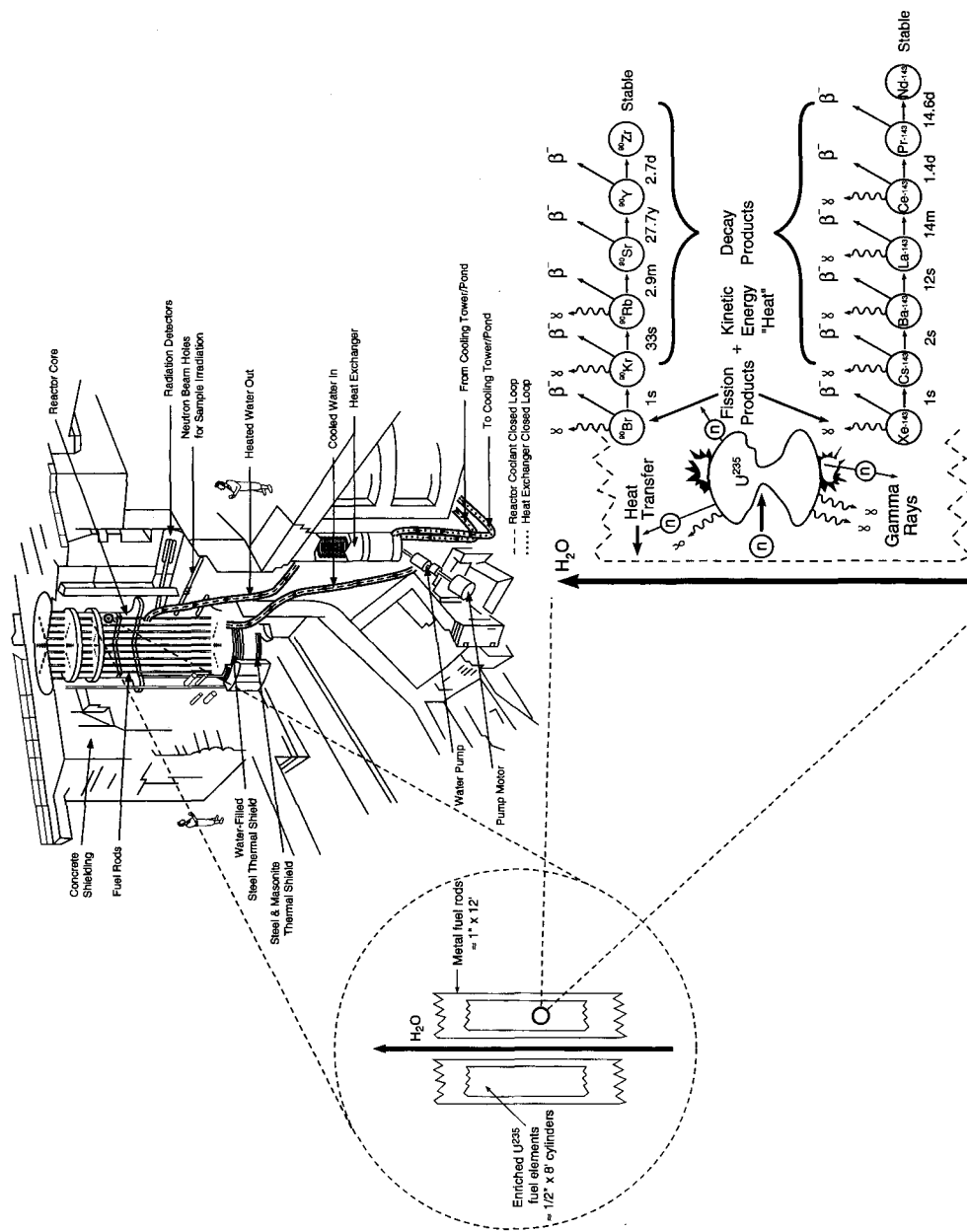


FIGURE 19-6. NRU Radionuclide research/production reactor (courtesy of Atomic Energy of Canada and Chalk River Laboratories, Chalk River, Ontario). Fuel rod assemblies and the fission process are illustrated to show some of the detail and the relationships associated with fission-produced radionuclides.

chemically separated from other fission products with essentially no stable isotopes (*carrier*) of the radionuclide present. Thus, the concentration or specific activity (measured in curies per gram) of these “carrier-free” fission-produced radionuclides is very high. High-specific-activity, carrier-free nuclides are preferred in radiopharmaceutical preparations to increase the labeling efficiency of the preparations and minimize the volume of the injected dose.

Neutron Activation–Produced Radionuclides

Neutrons produced by the fission of uranium in a nuclear reactor can be used to create radionuclides by bombarding stable target material placed in the reactor. This process, called *neutron activation*, involves the capture of neutrons by stable nuclei, which results in the production of radioactive nuclei. The most common neutron capture reaction for thermal (slow) neutrons is the (n,γ) reaction, in which the capture of the neutron by a nucleus is immediately followed by the emission of a gamma ray. Other thermal neutron capture reactions include the (n,p) and (n,α) reactions, in which the neutron capture is followed by the emission of a proton or an alpha particle, respectively. However, because thermal neutrons can induce these reactions only in a few, low-atomic-mass target nuclides, most neutron activation uses the (n,γ) reaction. Almost all radionuclides produced by neutron activation decay by beta-minus particle emission. Examples of radionuclides produced by neutron activation useful to nuclear medicine are listed below.

Phosphorus 32 production: $^{31}\text{P} \text{ (n,}\gamma\text{) } ^{32}\text{P}$ ($T_{1/2} = 14.3 \text{ days}$) [19-5]
 Chromium 51 production: $^{50}\text{Cr} \text{ (n,}\gamma\text{) } ^{51}\text{Cr}$ ($T_{1/2} = 27.8 \text{ days}$)

A radionuclide produced by an (n,γ) reaction is an isotope of the target element. As such, its chemistry is identical to that of the target material, making chemical separation techniques useless. Furthermore, no matter how long the target material is irradiated by neutrons, only a fraction of the target atoms will undergo neutron capture and become activated. Therefore, the material removed from the reactor will not be carrier free because it will always contain stable isotopes of the radionuclide. In addition, impurities in the target material will cause the production of other activated radionuclidic elements. The presence of carrier in the mixture limits the ability to concentrate the radionuclide of interest and therefore lowers the specific activity. For this reason, many of the clinically used radionuclides that could be produced by neutron activation (e.g., ^{133}I , ^{99}Mo) are instead produced by nuclear fission to maximize specific activity. An exception to the limitations of neutron activation is the production of ^{125}I , in which neutron activation of the target material, ^{124}Xe , produces a radioisotope, ^{125}Xe , that decays to form the desired radioisotope (Equation 19-6). In this case, the product radioisotope can be chemically or physically separated from the target material. Various characteristics of radionuclide production are compared in Table 19-1.

Iodine 125 production:

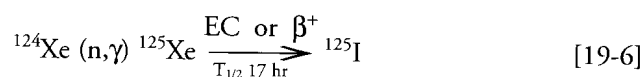


TABLE 19-1. COMPARISON OF RADIONUCLIDE PRODUCTION METHODS

Characteristic	Production method			
	Cyclotron	Nuclear reactor (fission)	Nuclear reactor (neutron activation)	Radionuclide generator
Bombarding particle	Proton, deuteron, triton, alpha	Neutron	Neutron	Production by decay of parent
Product	Neutron poor	Neutron excess	Neutron excess	Neutron poor or excess
Typical decay pathway	Positron emission, electron capture	Beta-minus	Beta-minus	Several modes
Typically carrier free	Yes	Yes	No	Yes
High specific activity	Yes	Yes	No	Yes
Relative cost	High	Low	Low	Low (^{99m}Tc) High (^{81m}Kr)
Radionuclides for nuclear medicine applications	^{201}Tl , ^{123}I , ^{67}Ga , ^{111}In , ^{18}F , ^{15}O , ^{57}Co	^{99}Mo , ^{131}I , ^{133}Xe	^{32}P , ^{51}Cr , ^{125}I , ^{89}Sr , ^{153}Sm	^{99m}Tc , ^{81m}Kr , ^{68}Ga , ^{82}Rb

Radionuclide Generators

Since the mid-1960s, technetium-99m (Tc-99m) has been the most important radionuclide used in nuclear medicine for a wide variety of radiopharmaceutical applications. However, its relatively short half-life (6 hours) makes it impractical to store even a weekly supply. This supply problem is overcome by obtaining the parent Mo-99, which has a longer half-life (67 hours) and continually produces Tc-99m. The Tc-99m is collected periodically in sufficient quantities for clinical operations. A system for holding the parent in such a way that the daughter can be easily separated for clinical use is called a *radionuclide generator*.

Molybdenum 99/Technetium-99m Radionuclide Generator

Technetium-99m pertechnetate ($^{99m}\text{TcO}_4^-$) is produced in a sterile, pyrogen-free form with high specific activity and a pH (~5.5) that is ideally suited for radiopharmaceutical preparations. The Mo-99 (produced by nuclear fission of U-235 to yield a high-specific-activity, carrier-free parent) is loaded, in the form of ammonium molybdenate ($(\text{NH}_4^+)(\text{MoO}_4^-)$), onto a porous column containing 5 to 10 g of an inorganic alumina (Al_2O_3) resin. The ammonium molybdenate becomes attached to the surface of the alumina molecules (a process called *adsorption*). The porous nature of the alumina provides a large surface area for adsorption of the parent.

As with all radionuclide generators, the chemical properties of the parent and daughter are different. In the Mo-99/Tc-99m or “moly” generator, the Tc-99m is much less tightly bound than the Mo-99. The daughter is removed (*eluted*) by the flow of isotonic (normal, 0.9%) saline (the “eluant”) through the column.

Two types of moly generators are commercially available. A “wet generator” has a large reservoir of saline connected by tubing to one end of the column; an evacuated vial draws the saline through the column, thus keeping the column wet. Figure 19.7 is a picture and cross-sectional diagram of a wet moly generator together with an insert

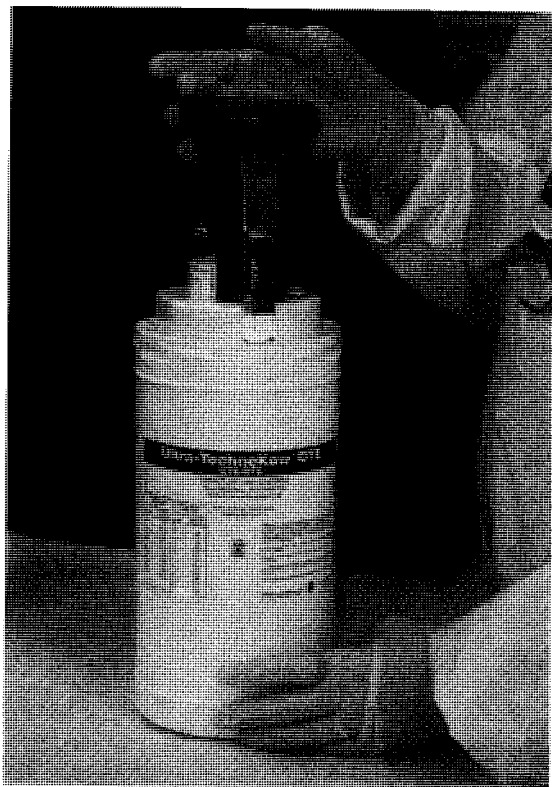
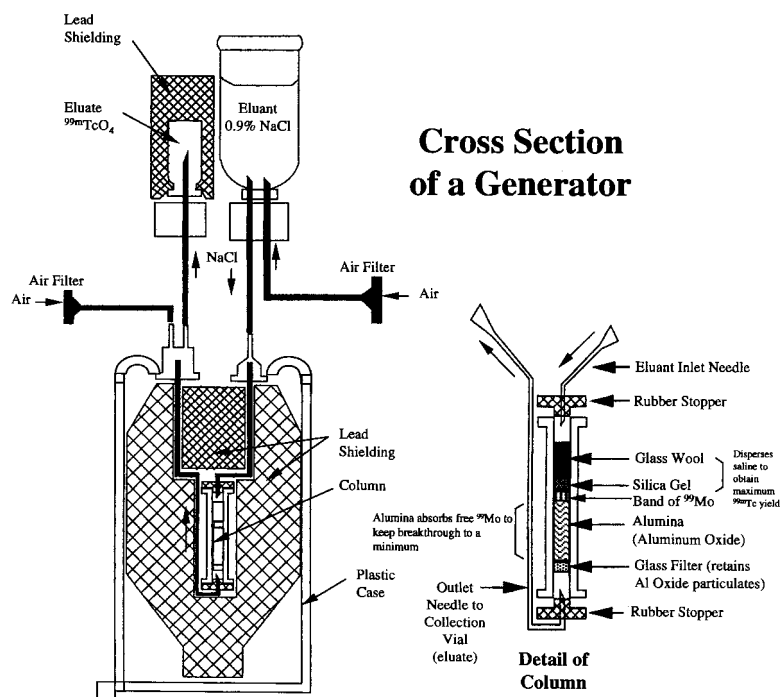


FIGURE 19-7. Picture of a "wet" molybdenum 99/technetium 99m generator and cross-sectional diagram of the generator. (Adapted with permission of Mallinckrodt Medical Inc., St. Louis, MO.)



that shows details of the generator column. A “dry generator” requires a small vial containing normal saline to be attached to one end of a dry column and a vacuum extraction vial to the other. On insertion of the vacuum collection vial, saline is drawn through the column and the eluate is collected, thus keeping the column dry between elutions. Sterility is achieved by a millipore filter connected to the end of the column, by the use of a bacteriostatic agent in the eluant, or by autoclave sterilization of the loaded column by the manufacturer. When the saline solution is passed through the column, the chloride ions easily exchange with the TcO_4^- (but not the MoO_4^-) ions, producing sodium pertechnetate, $\text{Na}^+ (^{99\text{m}}\text{TcO}_4^-)$.

Moly generators are typically delivered with approximately 37 to 111 GBq (1 to 3 Ci) of Mo-99, depending on the work load of the department. The activity of the daughter at the time of elution depends on the following:

1. The activity of the parent
2. The rate of formation of the daughter, which is equal to the rate of decay of the parent (i.e., $A_0 e^{-\lambda_P t}$)
3. The decay rate of the daughter
4. The time since the last elution
5. The elution efficiency (typically 80% to 90%).

Transient Equilibrium

Between elutions, the daughter (Tc-99m) builds up or “grows in” as the parent (Mo-99) continues to decay. After approximately 23 hours the Tc-99m activity reaches a maximum, at which time the production rate and the decay rate are equal and the parent and daughter are said to be in *transient equilibrium*. Once transient equilibrium has

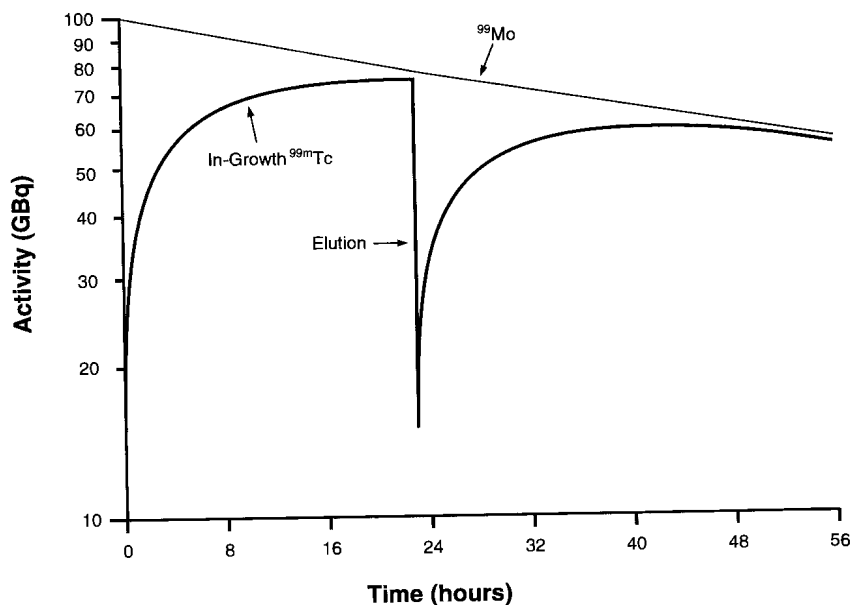


FIGURE 19-8. Time-activity curve of a molybdenum 99/technetium 99m radionuclide generator system demonstrating the in-growth of Tc-99m and subsequent elution.

been achieved, the daughter activity decreases, with an apparent half-life equal to the half-life of the parent. Transient equilibrium occurs when the half-life of the parent is greater than that of the daughter by a factor of approximately 10. When all the parent decays to a single daughter nuclide, the activity of the daughter will slightly exceed that of the parent at equilibrium. However, a fraction of Mo-99 decays directly to Tc-99 without first producing Tc-99m. Therefore, at equilibrium, the Tc-99m activity will be only approximately 96% that of the parent (Mo-99) activity.

Moly generators (sometimes called “cows”) are usually delivered weekly and eluted (called “milking the cow”) each morning, allowing for maximal in-growth of the daughter. Commercial moly generators contain between 9.25 and 111 GBq (between 0.25 and 3 Ci) of Mo-99 at the time of delivery. The elution process is approximately 90% efficient. This fact, together with the limitations on Tc-99m yield in the Mo-99 decay scheme, result in a maximum elution yield of approximately 85% of the Mo-99 activity at the time of elution. Therefore, a typical elution on Monday morning from a moly generator with 55.5 GBq (1.5 Ci) of Mo-99 yields approximately 47.2 GBq (1.28 Ci) of Tc-99m in 10 mL of normal saline (a common elution volume). By Friday morning of that week, the same generator would be capable of delivering only about 17.2 GBq (0.47 Ci). The moly generator can be eluted more frequently than every 23 hours; however, the Tc-99m yield will be less. Approximately half of the maximal yield will be available 6 hours after the last elution. Figure 19-8 shows a typical time-activity curve for a moly generator with an elution at 23 hours.

Secular Equilibrium

Although the moly generator is by far the most widely used in nuclear medicine, other generator systems produce clinically useful radionuclides. The characteristics of radionuclide generator systems are compared in Table 19-2. When the half-life

TABLE 19-2. CLINICALLY USED RADIONUCLIDE GENERATOR SYSTEMS IN NUCLEAR MEDICINE

Parent	Decay mode → Half-life	Daughter	Time of maximal ingrowth (equilibrium)	Decay mode → Half-life	Decay product
Germanium 69 (⁶⁹ Ge)	EC → 271 days	Gallium 68 (⁶⁸ Ga)	~6.5 hr (S)	β ⁺ , EC → 68 min	Zinc 68 (⁶⁸ Zn), stable
Rubidium 81 (⁸¹ Rb)	β ⁺ , EC → 4.5 hr	Krypton 81m (^{81m} Kr)	~80 sec (S)	IT → 13.5 sec	Krypton 81 ⁸¹ Kr ^a
Strontium 82 (⁸² St)	EC → 25.5 days	Rubidium 82 (⁸² Rb)	~7.5 min (S)	β ⁺ → 75 sec	Krypton 82 (⁸² Kr), stable
Molybdenum 99 (⁹⁹ Mo)	β ⁻ → 67 hr	Technetium 99m (^{99m} Tc)	~24 hr (T)	IT → 6 hr	Technetium 99 (⁹⁹ Tc) ^a

Note: Decay modes: EC, electron capture; β⁺, positron emission; β⁻, beta-minus; IT, isometric transition (i.e., gamma ray emission). Radionuclide equilibrium; T, transient; S, secular.

^aThese nuclides have half-lives greater than 10⁵ yr and for medical applications can be considered to be essentially stable.

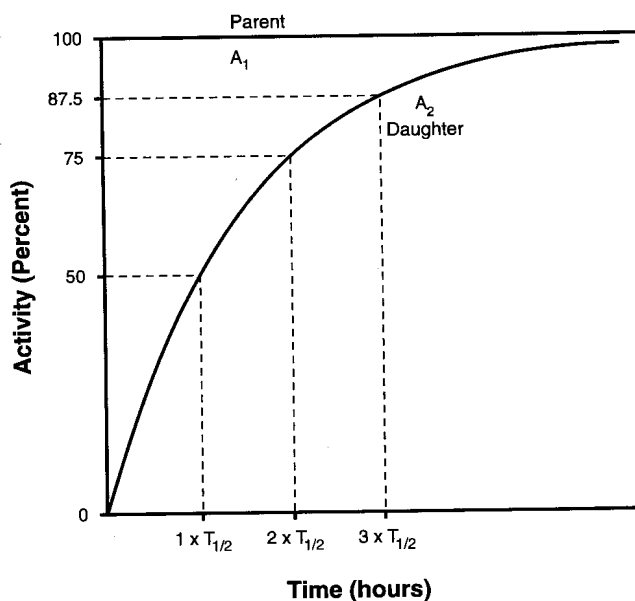


FIGURE 19-9. Time-activity curve demonstrating secular equilibrium.

of the parent is much longer than that of the daughter (i.e., more than about 100× longer), *secular equilibrium* occurs after approximately five to six half-lives of the daughter. In secular equilibrium, the activity of the parent and the daughter are the same if all of the parent atoms decay directly to the daughter. Once secular equilibrium is achieved, the daughter will have an apparent half-life equal to that of the parent. The rubidium 81/krypton 81m (Rb-81/Kr-81m) generator, with parent and daughter half-lives of 4.5 hours and 13.5 seconds, respectively, reaches secular equilibrium within approximately 1 minute after elution. Figure 19-9 shows a time-activity curve demonstrating secular equilibrium.

Quality Control

The users of moly generators are required to perform molybdenum and alumina breakthrough tests. Mo-99 contamination in the Tc-99m eluate is called *molybdenum breakthrough*. Mo-99 is an undesirable contaminant because its long half-life and highly energetic betas increase the radiation dose to the patient without providing any clinical information. The high-energy gamma rays (~740 and 780 keV) are very penetrating and cannot be efficiently detected by scintillation cameras. The *U.S. Pharmacopeia* (USP) and the U.S. Nuclear Regulatory Commission (NRC) limit the Mo-99 contamination to 0.15 μCi of Mo-99 per 1 mCi of Tc-99m at the time of administration. The Mo-99 contamination is evaluated by placing the Tc-99m eluate in a thick (~6 mm) lead container (provided by the dose calibrator manufacturer), which is placed in the dose calibrator. The high-energy photons of Mo-99 can be detected, whereas virtually all of the Tc-99m 140 keV photons are attenuated by the lead container. Eluates from moly generators rarely exceed permissible Mo-99 contamination limits. It is also possible (although rare) for some of the alumina from the column to contaminate the Tc-99m eluate. Alumina inter-

feres with the preparation of some of the radiopharmaceutical compounds (especially sulfur colloid and Tc-99m-labeled red blood cell [RBC] preparations). The USP limits the amount of alumina to no more than 10 μg alumina per 1 mL of Tc-99m eluate. Commercially available paper test strips and test standards are used to assay for alumina concentrations.

19.2 RADIOPHARMACEUTICALS

Characteristics, Applications, Quality Control, and Regulatory Issues in Medical Imaging

The vast majority of radiopharmaceuticals in nuclear medicine today use Tc-99m as the radionuclide. Most Tc-99m radiopharmaceuticals are easily prepared by aseptically injecting a known quantity of Tc-99m pertechnetate into a sterile vial containing the lyophilized (freeze-dried) pharmaceutical. The radiopharmaceutical complex is, in most cases, formed instantaneously and can be used for multiple doses over a period of several hours. Radiopharmaceuticals can be prepared in this fashion (called “kits”) as needed in the nuclear medicine department, or they may be delivered to the department by a centralized commercial radiopharmacy that serves several hospitals in the area. Although most Tc-99m radiopharmaceuticals can be prepared rapidly and easily at room temperature, several products (e.g., Tc-99m macroaggregated albumin [MAA]), require multiple steps such as boiling the Tc-99m reagent complex for several minutes. In almost all cases, however, the procedures are simple and the labeling efficiencies are very high (typically greater than 95%).

Other radionuclides common to diagnostic nuclear medicine imaging include ^{123}I , ^{67}Ga , ^{111}In , ^{133}Xe , and ^{201}Tl . Positron-emitting radionuclides are used for PET. ^{18}F , as fluorodeoxyglucose (FDG), is used in approximately 85% of all clinical PET applications. A wide variety of other positron-emitting radionuclides are currently being evaluated for their clinical utility, including carbon 11 (^{11}C), nitrogen 13 (^{13}N), oxygen 15 (^{15}O), gallium 68 (^{68}Ga), and rubidium 82 (^{82}Rb). The physical characteristics, most common modes of production, decay characteristics, and primary imaging photons (where applicable) of the radionuclides used in nuclear medicine are summarized in Table 19-3.

Ideal Radiopharmaceuticals

Although there are no “ideal” radiopharmaceuticals, it is helpful to think of currently used agents in light of the ideal characteristics.

Low Radiation Dose

It is important to minimize the radiation exposure to patients while preserving the diagnostic quality of the image. Radionuclides can be selected that have few particulate emissions and a high abundance of clinically useful photons. Most modern scintillation cameras are optimized for photon energies close to 140 keV, which is a compromise among patient attenuation, spatial resolution, and detection efficiency. Photons whose energies are too low are largely attenuated by the body, increasing the patient dose without contributing to image formation. High-energy photons escape the body but have poor detection efficiency and easily penetrate collimator

TABLE 19-3. PHYSICAL CHARACTERISTICS OF CLINICALLY USED RADIONUCLIDES

Radionuclide	Method of production	Mode of decay (%)	Principal photons used keV (% abundance)	Physical half-life	Comments
Chromium-51 (^{51}Cr)	Neutron activation	EC (100)	320 (9)	27.8 days	Used for in vivo red cell mass determinations (not used for imaging; samples counted in a NaI(Tl) well counter).
Cobalt-57 (^{57}Co)	Cyclotron produced	EC (100)	122 (86) 136 (11)	271 days	Principally used as a uniform flood field source for scintillation camera quality control.
Gallium-67 (^{67}Ga)	Cyclotron produced	EC (100)	93 (40) 184 (20) 300 (17) 393 (4)	78 hr	Typically use the 93, 184, and 300 keV photons for imaging.
Indium-111 (^{111}In)	Cyclotron produced	EC (100)	171 (90) 245 (94)	2.8 days	Typically used when the kinetics require imaging more than 24 hr after injection. Both photons are used in imaging.
Iodine-123 (^{123}I)	Cyclotron produced	EC (100)	159 (83)	13.2 hr	Has replaced ^{131}I for diagnostic imaging to reduce patient radiation dose.
Iodine-125 (^{125}I)	Neutron activation	EC (100)	35 (6) 27 (39) 28 (76) 31 (20)	60.2 days	Typically used as ^{125}I albumin for in vivo blood/plasma volume determinations (not used for imaging; samples counted in a NaI(Tl) well counter).
Iodine-131 (^{131}I)	Nuclear fission (^{235}U)	β^- (100)	284 (6) 364 (81) 637 (7)	8.0 days	Used for therapy: 364-keV photon used for imaging. Resolution and detection efficiency are poor due to high energy of photons. High patient dose, mostly from β^- particles.
Krypton-81m ($^{81\text{m}}\text{Kr}$)	Generator product	IT (100)	190 (67)	13 sec	This ultrashort-lived generator-produced radionuclide is a gas and can be used to perform serial lung ventilation studies with very little radiation exposure to patient or staff. The expense and short $T_{1/2}$ of the parent (^{81}Rb , 4.6 hr) limits its use.

Molybdenum-99 (⁹⁹ Mo)	Nuclear fission (²³⁵ U)	β ⁻ (100)	181 (6) 740 (12) 780 (4) None 103 (29)	67 hr	Parent material for Mo/Tc generator. Not used directly as a radiopharmaceutical; 740- and 780-keV photons used to identify "moly breakthrough." Prepared as either sodium or chromic phosphate. Therapy: used for pain relief from metastatic bone lesions. Advantage compared to ⁸⁹ Sr is that the ¹⁵³ Sm distribution can be imaged. Therapy: used for pain relief from metastatic bone lesions.
Phosphorus-32 (³² P)	Neutron activation	β ⁻ (100)		14.26 days	
Samarium-153 (¹⁵³ Sm)	Neutron activation	β ⁻ (100)		1.93 days	
Strontium-89 (⁸⁹ Sr)	Neutron activation	β ⁻ (100)	None	50.53 days	
Technetium-99m (^{99m} Tc)	Generator product	IT (100)	140 (88)	6.02 hr	This radionuclide accounts for more than 70% of all imaging studies.
Xenon-133 (¹³³ Xe)	Nuclear fission (²³⁵ U)	β ⁻ (100)	81 (37)	5.3 days	¹³³ Xe is a heavier-than-air gas. Low abundance and low energy of photon reduces image resolution.
Thallium-201 (²⁰¹ Tl)	Cyclotron produced	EC (100)	69–80 (94) 167 (10)	73.1 hr	The majority of clinically useful photons are low-energy x-rays (69–80 keV) from mercury 201 (²⁰¹ Hg), the daughter of ²⁰¹ Tl. Although these photons are in high abundance (94%), their low energy results in significant patient attenuation, which is particularly difficult with female patients, in whom breast artifacts in myocardial imaging often makes interpretation more difficult.
Positron-emitting radionuclides					
Carbon-11 (¹¹ C)	Cyclotron produced	β ⁺ (100)	511 (AR)	20.4 min	This radionuclide accounts for more than 80% of all clinical positron emission tomography studies; typically formulated as fluorodeoxyglucose (FDG).
Fluorine-18 (¹⁸ F)	Cyclotron produced	β ⁺ (97) EC (3)	511 (AR)	110 min	
Nitrogen-13 (¹³ N)	Cyclotron produced	β ⁺ (100)	511 (AR)	10 min	
Gallium-68 (⁶⁸ Ga)	Cyclotron produced	β ⁺ (89) EC (11)	511 (AR)	68 min	
Rubidium-82 (⁸² Rb)	Generator product	β ⁺ (100)	511 (AR)	75 sec	

Note: β⁻, Beta-minus decay; β⁺, beta-plus (positron) decay; AR, annihilation radiation; EC, electron capture; IT, isomeric transition (i.e., gamma ray emission).

septa (see Chapter 21). A radiopharmaceutical should have an effective half-life long enough to complete the study with an adequate concentration in the organ of interest but short enough to minimize the patient dose.

High Target/Nontarget Activity

The ability to detect and evaluate lesions depends largely on the concentration of the radiopharmaceutical in the organ or lesion of interest or on a clinically useful uptake and clearance pattern. To maximize the concentration of the radiopharmaceutical in the organ or lesion of interest and thus the image contrast, there should be little accumulation in other tissue and organs. Radiopharmaceuticals are developed to maximize target/nontarget ratio for various organ systems. Abnormalities can be identified as localized areas of increased radiopharmaceutical concentration, called “hot spots” (e.g., a stress fracture in a bone scan), or as “cold spots” in which the radiopharmaceutical’s normal localization in a tissue is altered by a disease process (e.g., perfusion defect in a lung scan with ^{99m}Tc -MAA). Disassociation of the radionuclide from the radiopharmaceutical alters the desired biodistribution, thus degrading image quality. Good quality control over radiopharmaceutical preparation helps to ensure that the radionuclide is bound to the pharmaceutical.

Safety, Convenience, and Cost-Effectiveness

Low toxicity is enhanced by the use of high-specific-activity, carrier-free radionuclides that also facilitate radiopharmaceutical preparation and minimize the required amount of the isotope. For example, 3.7 GBq (100 mCi) of I-131 contains only 0.833 μg of iodine. Radionuclides should also have a chemical form, pH, concentration, and other characteristics that facilitate rapid complexing with the pharmaceutical under normal laboratory conditions. The compounded radiopharmaceutical should be stable, with a shelf life compatible with clinical use, and should be readily available from several manufacturers to minimize cost.

Radiopharmaceutical Mechanisms of Localization

Radiopharmaceutical concentration in tissue is driven by one or more of the following mechanisms: (a) compartmental localization and leakage, (b) cell sequestration, (c) phagocytosis, (d) passive diffusion, (e) metabolism, (f) active transport, (g) capillary blockade, (h) perfusion, (i) chemotaxis, (j) antibody-antigen complexation, (k) receptor binding, and (l) physiochemical adsorption.

Compartmental Localization and Leakage

Compartmental localization refers to the introduction of the radiopharmaceutical into a well-defined anatomic compartment. Examples include Xe-133 gas inhalation into the lung, intraperitoneal instillation of P-32 chromic phosphate, and Tc-99m-labeled RBCs injected into the circulatory system. Compartmental leakage is used to identify an abnormal opening in an otherwise closed compartment, as when labeled RBCs are used to detect gastrointestinal bleeding.

Cell Sequestration

RBCs are withdrawn from the patient, labeled with Tc-99m, and slightly damaged by *in vitro* heating in a boiling water bath for approximately 30 minutes. After they have been reinjected, the spleen's ability to recognize and remove (i.e., sequester) the damaged RBCs is evaluated. This procedure allows for the evaluation of both splenic morphology and function.

Phagocytosis

The cells of reticuloendothelial system are distributed in the liver (~85%), spleen (~10%), and bone marrow (~5%). These cells recognize small foreign substances in the blood and remove them by phagocytosis. In a liver scan, for example, the Tc-99m labeled sulfur colloid particles (~100 nm) are recognized, being substantially smaller than circulating cellular elements, and are rapidly removed from circulation.

Passive Diffusion

Passive diffusion is simply the free movement of a substance from a region of high concentration to one of lower concentration. Anatomic and physiologic mechanisms exist in the brain tissue and surrounding vasculature that allow essential nutrients, metabolites, and lipid-soluble compounds to pass freely between the plasma and brain tissue while many water-soluble substances (including most radiopharmaceuticals) are prevented from entering healthy brain tissue. This system, called the blood-brain barrier, protects and regulates access to the brain. Disruptions of the blood-brain barrier can be produced by trauma, neoplasms, and inflammatory changes. The disruption permits radiopharmaceuticals such as Tc-99m-diethylenetriaminepentaacetic acid (DTPA), which is normally excluded by the blood-brain barrier, to follow the concentration gradient and enter the affected brain tissue.

Metabolism

FDG is a glucose analogue: its increased uptake correlates with increased glucose metabolism. FDG crosses the blood-brain barrier, where it is metabolized by brain cells. FDG enters cells via carrier-mediated diffusion (as does glucose) and is then phosphorylated and trapped in brain tissue for several hours as FDG-6-phosphate.

Active Transport

Active transport involves cellular metabolic processes that expend energy to concentrate the radiopharmaceutical into a tissue against a concentration gradient above plasma levels. The classic example in nuclear medicine is the trapping and organification of radioactive iodide. Trapping of iodide in the thyroid gland occurs against a concentration gradient where it is oxidized to a highly reactive iodine by peroxidase enzyme. Organification follows, resulting in the production of radiolabeled triiodothyronine (T_3) and thyroxine (T_4). Another example is the localization of thallium (a potassium analog) in muscle tissue. The concentration of Tl-201 is

mediated by the energy-dependent Na^+/K^+ ionic pump. Nonuniform distribution of Tl-201 in the myocardium indicates a myocardial perfusion deficit.

Capillary Blockade

When particles slightly larger than RBCs are injected intravenously, they become trapped in the narrow capillary beds. A common example in nuclear medicine is in the assessment of pulmonary perfusion by the injection of Tc-99m-MAA, which is trapped in the pulmonary capillary bed. Imaging the distribution of Tc-99m-MAA provides a representative assessment of pulmonary perfusion. The “microemboli” created by this radiopharmaceutical do not pose a significant clinical risk because only a very small percentage of the pulmonary capillaries are blocked and the MAA is eventually removed by biodegradation.

Perfusion

Relative perfusion of a tissue or organ system is an important diagnostic element in many nuclear medical procedures. For example, the perfusion phase of a three-phase bone scan helps to distinguish between an acute process (e.g., osteomyelitis) and remote fracture. Perfusion is also an important diagnostic element in examinations such as renograms and cerebral and hepatic blood flow studies.

Chemotaxis

Chemotaxis describes the movement of a cell such as a leukocyte in response to a chemical stimulus. ^{111}In -labeled leukocytes respond to products formed in immunologic reactions by migrating and accumulating at the site of the reaction as part of an overall inflammatory response.

Antibody-Antigen Complexation

An antigen is a biomolecule (typically a protein) that is capable of inducing the production of, and binding to, an antibody in the body. The antigen has a strong and specific affinity for the antibody. An *in vitro* test (developed by Berson and Yalow in the late 1950s) called radioimmunoassay (RIA) makes use of the competition between a radiolabeled antigen and the same antigen in the patient's serum for antibody binding sites. Although it was originally developed as a serum insulin assay, today RIA techniques are employed extensively to measure minute quantities of various enzymes, antigens, drugs, and hormones. At equilibrium, the more unlabeled serum antigen that is present, the less radiolabeled antigen (free antigen) will become bound to the antibody (bound antigen). The serum level is measured by comparing the ratio between bound and free antigen in the sample to a known standard for that particular assay.

Antigen-antibody complexation is also used in diagnostic imaging with such agents as In-111-labeled antibodies for the detection of colorectal carcinoma. This class of immunospecific radiopharmaceuticals promises to provide an exciting new approach to diagnostic imaging. In addition, a variety of radiolabeled (typically I-131-labeled) monoclonal antibodies directed toward tumors are being used in an attempt to deliver tumoricidal radiation doses. This procedure, called radioimmunotherapy, is under clinical investigation as an adjunct to conventional cancer therapy.

Receptor Binding

This class of radiopharmaceuticals is characterized by their high affinity to bind to specific receptor sites. For example, the uptake of In-111-octreotide, used for the localization of neuroendocrine and other tumors, is based on the binding of a somatostatin analog to receptor sites in tumors.

Physiochemical Adsorption

Methylenediphosphonate (MDP) localization occurs primarily by adsorption in the mineral phase of the bone. MDP concentrations are significantly higher in amorphous calcium than in mature hydroxyapatite crystalline structures, which helps to explain its concentration in areas of increased osteogenic activity.

A summary of the characteristics and clinical utility of common radiopharmaceuticals is provided in Appendix D.

Radiopharmaceutical Quality Control

Aside from the radionuclidic purity quality control performed on the ^{99m}Tc -pertechnetate generator eluate, the most common radiopharmaceutical quality control procedure is the test for radiochemical purity. The radiochemical purity assay identifies the fraction of the total radioactivity that is in the desired chemical form. Radiochemical impurity can occur as the result of temperature changes, presence of unwanted oxidizing or reducing agents, pH changes, or radiation damage to the radiopharmaceutical (called autoradiolysis). The presence of radiochemical impurities compromises the diagnostic utility of the radiopharmaceutical by reducing uptake in the organ of interest and increasing background activity, thereby degrading image quality. In addition to lowering the diagnostic quality of the examination, radiochemical impurities unnecessarily increase patient dose.

The most common method to determine the amount of radiochemical impurity in a radiopharmaceutical preparation is thin-layer chromatography. This test is performed by placing a small aliquot (~1 drop) of the radiopharmaceutical preparation approximately 1 cm from one end of a small rectangular paper (e.g., Whatman filter paper) or a silica-coated plastic strip. This strip is called the "stationary phase." The spotted end of the strip is then lowered into a glass vial containing an appropriate solvent (e.g., saline, acetone, or 85% methanol), such that the solvent front begins just below the spot. The depth of the solvent in the vial must be low enough so that the spot of radiopharmaceutical on the strip is above the solvent. The solvent will slowly move up the strip, and the various radiochemicals will partition themselves at specific locations (identified as "reference factors," or R_f from 0 to 1) along the strip according to their relative solubilities. Once the solvent has reached a predetermined line near the top of the strip, the strip is removed and dried. The strip is cut into sections, and the percentage of the total radioactivity on each section of the strip is assayed and recorded. The movements of radiopharmaceuticals and their contaminants have been characterized for several solvents. Comparison of the results with these reference values allows the identity and percentage of the radiochemical impurities to be determined.

The two principal radiochemical impurities in technetium-labeled radiopharmaceuticals are free (i.e., unbound) Tc-99m-pertechnetate and hydrolyzed Tc-99m. The Tc-99m radiopharmaceutical complex and its associated impurities will,

depending on the solvent, either remain at the origin ($R_f = 0$) or move with the solvent front to the end of the strip ($R_f = 1$). For example, with a silica stationary phase in an acetone solvent, Tc-99m-MAA remains at the origin and any free pertechnetate or hydrolyzed Tc-99m moves with the solvent front. On the other hand, I-131 (as bound NaI) moves with the solvent front ($R_f = 1$) in an 85% methanol solvent, while the impurity (unbound iodine) moves ~20% of the distance from the origin ($R_f = 0.2$). These assays are easy to perform and should be used as part of a routine radiopharmacy quality control program and whenever there is a question about the radiochemical integrity of a radiopharmaceutical preparation.

19.3 REGULATORY ISSUES

Investigational Radiopharmaceuticals

All pharmaceuticals for human use, whether radioactive or not, are regulated by the U.S. Food and Drug Administration (FDA). A request to evaluate a new radiopharmaceutical for human use is submitted to the FDA in an application called a "Notice of Claimed Investigational Exemption for a New Drug" (IND). The IND can be sponsored by either an individual physician or a radiopharmaceutical manufacturer who will work with a group of clinical investigators to collect the necessary data. The IND application includes the names and credentials of the investigators, the clinical protocol, details of the research project, details of the manufacturing of the drug, and animal toxicology data. The clinical investigation of the new radiopharmaceutical occurs in three stages. Phase I focuses on a limited number of patients and is designed to provide information on the pharmacologic distribution, metabolism, dosimetry, toxicity, optimal dose schedule, and adverse reactions. Phase II studies include a limited number of patients with specific diseases to begin the assessment of the drug's efficacy, refine the dose schedule, and collect more information on safety. Phase III clinical trials involve a much larger number of patients (and are typically conducted by several institutions) to provide more extensive (i.e., statistically significant) information on efficacy, safety, and dose administration. To obtain approval to market a new radiopharmaceutical, a "New Drug Application" (NDA) must be submitted to and approved by the FDA. Approval of a new radiopharmaceutical typically requires 5 to 10 years from laboratory work to NDA. The package insert of an approved radiopharmaceutical describes the intended purpose of the radiopharmaceutical, the suggested dose, dosimetry, adverse reactions, clinical pharmacology, and contraindications.

Regulatory agencies require that all research involving human subjects be reviewed and approved by an institutional review board (IRB) and that informed consent be obtained from each research subject. Most academic institutions have IRBs. An IRB comprises clinical, scientific, legal, and other experts and must include at least one member who is not otherwise affiliated with the institution. Informed consent must be sought in a manner that minimizes the possibility of coercion and provides the subject sufficient opportunity to decide whether or not to participate. The information presented must be in language understandable to the subject. It must include a statement that the study involves research, the purposes of the research, a description of the procedures, and identification of any procedures that are experimental; a description of any reasonably foreseeable risks or discomforts; a description of any likely benefits to the subject or to others; a dis-

closure of alternative treatments; and a statement that participation is voluntary, that refusal will involve no penalty or loss of benefits, and that the subject may discontinue participation at any time.

Written Directives and Medical Events

Although the production of radiopharmaceuticals is regulated by the FDA, the medical use of byproduct material is regulated by the NRC or a comparable state agency. The NRC's regulations apply only to the use of byproduct material. Byproduct material means radionuclides that are the byproducts of nuclear fission (e.g., Tc-99m). Radionuclides produced by other methods, such as by cyclotrons (e.g., In-111), are not subject to NRC regulations, although they may be regulated by state regulatory agencies. The NRC's regulations regarding the medical use of byproduct material are contained in Title 10, Part 35, of the *Code of Federal Regulations* (10 CFR 35). The specific information presented on 10 CFR 35 is based on a final draft of a proposed revision to these regulations. The final regulations should be consulted when they are promulgated.

The NRC requires that, before the administration of a dosage of I-131 sodium iodide greater than 1.11 MBq (30 μ Ci) or any therapeutic dosage of unsealed byproduct material, a *written directive* must be prepared, signed, and dated by an authorized user. The written directive must contain the patient or human research subject's name and must describe the radioactive drug, the activity, and the route of administration. In addition, the NRC requires the implementation of written procedures to provide, for each administration requiring a written directive, high confidence that the patient or human research subject's identity is verified before the administration and that each administration is performed in accordance with the written directive.

The NRC defines certain errors in the administration of radiopharmaceuticals as *medical events*. Medical events were formerly called "misadministrations." Medical events must be reported to the NRC. The initial report must be made by telephone no later than the next calendar day after the discovery of the event and must be followed by a written report within 15 days. This report must include specific information such as a description of the incident, the cause of the medical event, the effect (if any) on the individual or individuals involved, and proposed corrective actions. The referring physician must be notified of the medical event, and the patient must also be notified, unless the referring physician states that he or she will inform the patient or that, based on medical judgment, notification of the patient would be harmful. Other reporting requirements can be found in 10 CFR 35.

The NRC defines a medical event as follows:

A. The administration of byproduct material that results in one of the following conditions (1 or 2) unless its occurrence was as the direct result of patient intervention (e.g., I-131 therapy patient takes only one-half of the prescribed treatment then refuses to take the balance of the prescribed dosage):

1. A dose that differs from the prescribed dose by more than 0.05 Sv (5 rem) effective dose equivalent, 0.5 Sv (50 rem) to an organ or tissue, or 0.5 Sv (50 rem) shallow dose equivalent to the skin; and one of the following conditions (*i* or *ii*) has also occurred.

(*i*) The total dose delivered differs from the prescribed dose by 20% or more;

(ii) The total dosage (i.e., administered activity) delivered differs from the prescribed dosage by 20% or more or falls outside the prescribed dosage range. Falling outside the prescribed dosage range means the administration of activity that is greater or less than a predetermined range of activity for a given procedure that has been established by the licensee (e.g., 370 to 1,110 MBq [10 to 30 mCi] ^{99m}Tc -MDP for an adult bone scan).

2. A dose that exceeds 0.05 Sv (5 rem) effective dose equivalent, 0.5 Sv (50 rem) to an organ or tissue, or 0.5 Sv (50 rem) shallow dose equivalent to the skin from any of the following:

- (i) An administration of a wrong radioactive drug containing byproduct material;
- (ii) An administration of a radioactive drug containing byproduct material by the wrong route of administration;
- (iii) An administration of a dose or dosage to the wrong individual or human research subject;

B. Any event resulting from intervention of a patient or human research subject in which the administration of byproduct material or radiation from byproduct material results or will result in unintended permanent functional damage to an organ or a physiological system, as determined by a physician. Patient intervention means actions by the patient or human research subject, whether intentional or unintentional, that affect the radiopharmaceutical administration.

This definition of a medical event was summarized from NRC regulations. It applies only to the use of unsealed byproduct material and omits the definition of a medical event involving the use of sealed sources of byproduct material to treat patients. The complete regulations regarding written directives and medical events can be found in 10 CFR 35. State regulatory requirements should be consulted, because they may differ from federal regulations.

SUGGESTED READING

- Hung JC, Ponto JA, Hammes RJ. Radiopharmaceutical-related pitfalls and artifacts. *Semin Nucl Med* 1996;26:208–255.
- Medley CM, Vivian GC. Radionuclide developments. *Br J Radiol* 1997;70:133–144.
- Ponto JA. The AAPM/RSNA physics tutorial for residents: radiopharmaceuticals. *Radiographics* 1998;18:1395–1404.
- Rhodes BA, Hladik WB, Norenberg JP. Clinical radiopharmacy: principles and practices. *Semin Nucl Med* 1996;26:77–84.
- Saha GB. *Fundamentals of nuclear pharmacy*, 4th ed. New York: Springer-Verlag, 1998.
- Sampson CB, ed. *Textbook of radiopharmacy: theory and practice*, 2nd ed. New York: Gordon and Breach Publishers, 1995.

RADIATION DETECTION AND MEASUREMENT

The detection and measurement of ionizing radiation is the basis for the majority of diagnostic imaging. In this chapter, the basic concepts of radiation detection and measurement are introduced, followed by a discussion of the characteristics of specific types of detectors. The electronic systems used for pulse height spectroscopy and the use of sodium iodide (NaI) scintillators to perform gamma-ray spectroscopy are described, followed by a discussion of detector applications. The use of radiation detectors in imaging devices is covered in other chapters.

All detectors of ionizing radiation require the interaction of the radiation with matter. Ionizing radiation deposits energy in matter by ionization and excitation. *Ionization* is the removal of electrons from atoms or molecules. (An atom or molecule stripped of an electron has a net positive charge and is called a *cation*. In many gases, the free electrons become attached to uncharged atoms or molecules, forming negatively charged *anions*. An ion pair consists of a cation and its associated free electron or anion.) *Excitation* is the elevation of electrons to excited states in atoms, molecules, or a crystal. Excitation and ionization may produce chemical changes or the emission of visible light. Most energy deposited by ionizing radiation is ultimately converted to heat.

The amount of energy deposited in matter by a single interaction is very small. For example, a 140-keV gamma ray deposits 2.24×10^{-14} joules if completely absorbed. Raising the temperature of 1 g of water 1°C (i.e., 1 calorie) would require the complete absorption of 187 trillion (187×10^{12}) of these photons. For this reason, most radiation detectors provide signal amplification. In detectors that produce an electrical signal, the amplification is electronic. In photographic film, the amplification is achieved chemically.

20.1 TYPES OF DETECTORS

Radiation detectors may be classified by their detection method. A *gas-filled detector* consists of a volume of gas between two electrodes. Ions produced in the gas by the radiation are collected by the electrodes, resulting in an electrical signal.

The interaction of ionizing radiation with certain materials produces ultraviolet and/or visible light. These materials are called *scintillators*. They are commonly attached to or incorporated in devices that convert the light into an electrical sig-

nal. For other applications, photographic film is used to record the light the scintillators emit. Many years ago, in physics research and medical fluoroscopy, the light from scintillators was viewed directly with dark-adapted eyes.

Semiconductor detectors are especially pure crystals of silicon, germanium, or other semiconductor materials to which trace amounts of impurity atoms have been added so that they act as diodes. A diode is an electronic device with two terminals that permits a large electrical current to flow when a voltage is applied in one direction, but very little current when the voltage is applied in the opposite direction. When used to detect radiation, a voltage is applied in the direction in which little current flows. When an interaction occurs in the crystal, electrons are raised to an excited state, allowing a momentary electrical current to flow through the device.

Detectors may also be classified by the type of information produced. Detectors, such as Geiger-Mueller (GM) detectors, that indicate the number of interactions occurring in the detector are called *counters*. Detectors that yield information about the energy distribution of the incident radiation, such as NaI scintillation detectors, are called *spectrometers*. Detectors that indicate the net amount of energy deposited in the detector by multiple interactions are called *dosimeters*.

Pulse and Current Modes of Operation

Many radiation detectors produce an electrical signal after each interaction of a particle or photon. The signal generated by the detector passes through a series of electronic circuits, each of which performs a function such as signal amplification, signal processing, or data storage. A detector and its associated electronic circuitry form a *detector system*. There are two fundamental ways that the circuitry may process the signal—pulse mode and current mode. In *pulse mode*, the signal from each interaction is processed individually. In *current mode*, the electrical signals from individual interactions are averaged together, forming a net current signal.

There are advantages and disadvantages to each method for handling the signal. GM detectors are operated in pulse mode, whereas most ionization chambers, including the dose calibrators used in nuclear medicine and those used in some computed tomography (CT) scanners, are operated in current mode. Scintillation detectors are operated in pulse mode in nuclear medicine applications but in current mode in digital radiography, fluoroscopy, and x-ray CT.

Effect of Interaction Rate on Detectors Operated in Pulse Mode

The main problem with using a radiation detector or detector system in pulse mode is that two interactions must be separated by a finite amount of time if they are to produce distinct signals. This interval is called the *dead time* of the system. If a second interaction occurs during this time interval, its signal will be lost; furthermore, if it is close enough in time to the first interaction, it may even distort the signal from the first interaction. The fraction of counts lost from dead-time effects tends to be small at low interaction rates and increases with interaction rate.

The dead time of a detector system is largely determined by the component in the series with the longest dead time. For example, the detector usually has the longest dead time in GM counter systems, whereas in multichannel analyzer systems (see later discussion) the analog-to-digital converter often has the longest dead time.

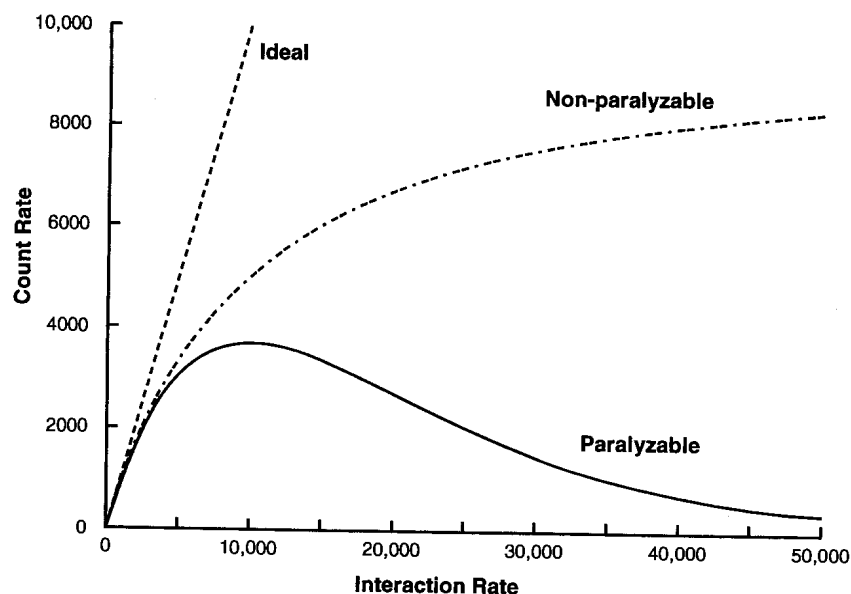


FIGURE 20-1. Effect of interaction rate on measured count rate of paralyzable and nonparalyzable detectors. The “ideal” line represents the response of a hypothetical detector that does not suffer from dead-time count losses (i.e., the count rate is equal to the interaction rate). Note that the y-axis scale is expanded with respect to that of the x-axis; the “ideal” line would be at a 45-degree angle if the scales were equal.

The dead times of different types of systems vary widely. GM counters have dead times ranging from tens to hundreds of microseconds, whereas most other systems have dead times of less than a few microseconds. It is important to know the count-rate behavior of a detector system; if a detector system is operated at too high an interaction rate, an artificially low count rate will be obtained.

There are two mathematical models describing the behavior of detector systems operated in pulse mode—paralyzable and nonparalyzable. Although these models are simplifications of the behavior of real detector systems, real systems often behave like one or the other model. In a *paralyzable* system, an interaction that occurs during the dead time after a previous interaction extends the dead time; in a *nonparalyzable* system, it does not. Figure 20-1 shows the count rates of paralyzable and nonparalyzable detector systems as a function of the rate of interactions in the detector. At very high interaction rates, a paralyzable system will be unable to detect any interactions after the first, because subsequent interactions will extend the dead time, causing the detector to indicate a count rate of zero!

Current Mode Operation

When a detector is operated in current mode, all information regarding individual interactions is lost. For example, neither the interaction rate nor the energies deposited by individual interactions can be determined. However, if the amount of electrical charge collected from each interaction is proportional to the energy deposited by that interaction, then the net electrical current is proportional to the dose rate in the detector material. Detectors subject to very high interaction rates

are often operated in current mode to avoid dead-time information losses. Ion chambers used in phototimed radiography, image-intensifier tubes in fluoroscopy, ion chambers or scintillation detectors in x-ray CT machines, and most nuclear medicine dose calibrators are all operated in current mode.

Spectroscopy

Spectroscopy is the study of the energy distribution of a radiation field, and a *spectrometer* is a detector that yields information about the energy distribution of the incident radiation. Most spectrometers are operated in pulse mode, and the amplitude of each pulse is proportional to the energy deposited in the detector by the interaction causing that pulse. *The energy deposited by an interaction, however, is not always the total energy of the incident particle or photon.* For example, a gamma ray may interact with the detector by Compton scattering, with the scattered photon escaping the detector. In this case, the deposited energy is the difference between the energies of the incident and scattered photons. A *pulse height spectrum* is usually depicted as a graph of the number of interactions depositing a particular amount of energy in the spectrometer as a function of energy (Fig. 20-2). Because the energy deposited by an interaction may be less than the total energy of the incident particle or photon and also because of statistical effects in the detector, *the pulse height spectrum produced by a spectrometer is not identical to the actual energy spectrum of the incident radiation.* The energy resolution of a spectrometer is a measure of its ability to differentiate between particles or photons of different energies. Pulse height spectroscopy is discussed later in this chapter.

Detection Efficiency

The *efficiency (sensitivity)* of a detector is a measure of its ability to detect radiation. The efficiency of a detection system operated in pulse mode is defined as the probability that a particle or photon emitted by a source will be detected. It is measured by placing a source of radiation in the vicinity of the detector and dividing the number of particles or photons detected by the number emitted:

$$\text{Efficiency} = \frac{\text{Number detected}}{\text{Number emitted}} \quad [20-1]$$

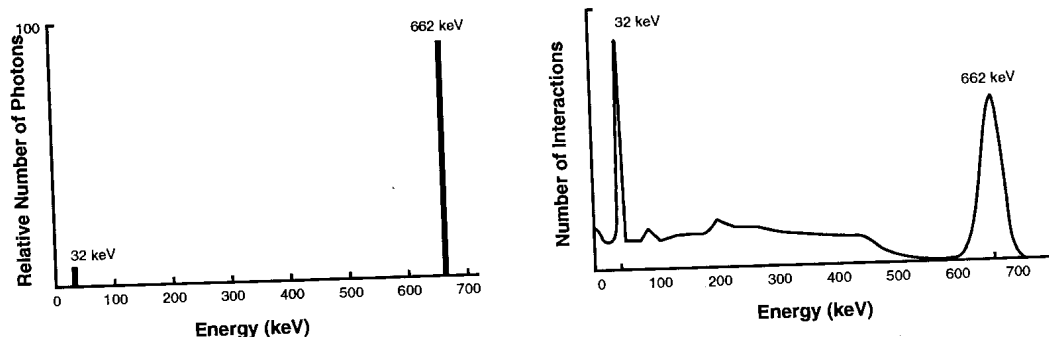


FIGURE 20-2. Energy spectrum of cesium 137 (left) and resultant pulse height spectrum from detector (right).

This equation can be written as follows:

$$\text{Efficiency} = \frac{\text{Number reaching detector}}{\text{Number emitted}} \times \frac{\text{Number detected}}{\text{Number reaching detector}}$$

Therefore, the detection efficiency is the product of two terms, the geometric efficiency and the intrinsic efficiency:

$$\text{Efficiency} = \text{Geometric efficiency} \times \text{Intrinsic efficiency} \quad [20-2]$$

where the *geometric efficiency* of a detector is the fraction of emitted particles or photons that reach the detector and the *intrinsic efficiency* is the fraction of those particles or photons reaching the detector that are detected. Because the total, geometric, and intrinsic efficiencies are all probabilities, each ranges from 0 to 1.

The geometric efficiency is determined by the geometric relationship between the source and the detector (Fig. 20-3). It increases as the source is moved toward the detector and approaches 0.5 when a point source is placed against a flat surface of the detector, because in that position one half of the photons or particles are emitted into the detector. For a source inside a well-type detector, the geometric efficiency approaches 1, because most of the particles or photons are intercepted by the detector. (A well-type detector is a detector containing a cavity for the insertion of samples.)

The intrinsic efficiency of a detector in detecting photons, often called the *quantum detection efficiency* or QDE, is determined by the energy of the photons and the atomic number, density, and thickness of the detector. If a parallel beam of mono-energetic photons is incident upon a detector of uniform thickness, the intrinsic efficiency of the detector is given by the following equation:

$$\text{Intrinsic efficiency} = 1 - e^{-\mu x} \quad [20-3]$$

where μ is the linear attenuation coefficient of the detector material and x is the thickness of the detector.

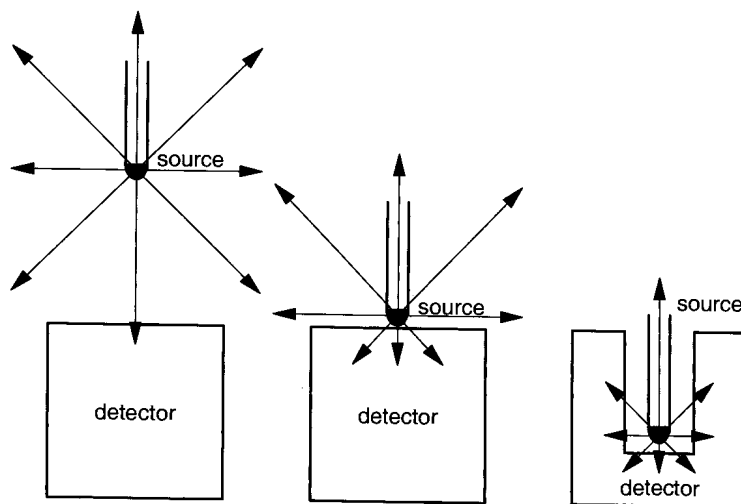


FIGURE 20-3. Geometric efficiency. With a source far from the detector (*left*), the geometric efficiency is less than 50%. With a source against the detector (*center*), the geometric efficiency is approximately 50%. With a source in a well detector (*right*), the geometric efficiency is greater than 50%.

20.2 GAS-FILLED DETECTORS

Basic Principles

A gas-filled detector (Fig. 20-4) consists of a volume of gas between two electrodes, with an electric potential difference (voltage) applied between the electrodes. Ionizing radiation forms ion pairs in the gas. The positive ions (cations) are attracted to the negative electrode (cathode), and the electrons or anions are attracted to the positive electrode (anode). In most detectors, the cathode is the wall of the container that holds the gas and the anode is a wire inside the container. After reaching the anode, the electrons travel through the circuit to the cathode, where they recombine with the cations. This electrical current can be measured with a sensitive ammeter or other electrical circuitry.

There are three types of gas-filled detectors in common use—ionization chambers, proportional counters, and GM counters. The type of detector is determined primarily by the voltage applied between the two electrodes. In ionization chambers, the two electrodes can have almost any configuration: they may be two parallel plates, two concentric cylinders, or a wire within a cylinder. In proportional counters and GM counters, the anode must be a thin wire. Figure 20-5 shows the amount of electrical charge collected after a single interaction as a function of the electrical potential difference (voltage) applied between the two electrodes.

Ionizing radiation produces ion pairs in the gas of the detector. If no voltage is applied between the electrodes, no current flows through the circuit because there is no electric field to attract the charged particles to the electrodes; the ion pairs merely recombine in the gas. When a small voltage is applied, some of the cations are attracted to the cathode and some of the electrons or anions are attracted to the anode before they can recombine. As the voltage is increased, more ions are collected and fewer recombine. This region, in which the current increases as the voltage is raised, is called the *recombination region* of the curve.

As the voltage is increased further, a plateau is reached in the curve. In this region, called the *ionization chamber region*, the applied electric field is sufficiently strong to collect almost all ion pairs; additional increases in the applied voltage do

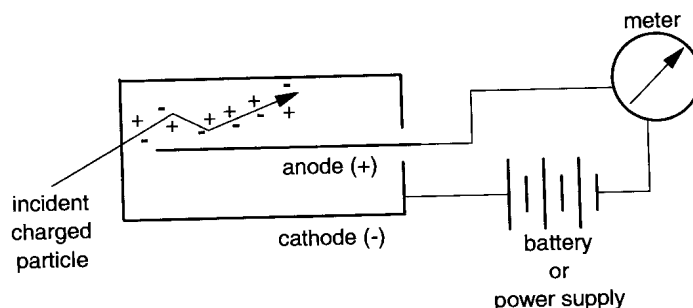


FIGURE 20-4. Gas-filled detector. A charged particle, such as a beta particle, is shown entering the tube from outside. This can occur only if the tube has a sufficiently thin wall. When a thick-wall gas-filled detector is used to detect x-rays and gamma rays, the charged particles causing the ionization are electrons generated by Compton and photoelectric interactions of the incident x-rays or gamma rays in the detector wall or in the gas in the detector.

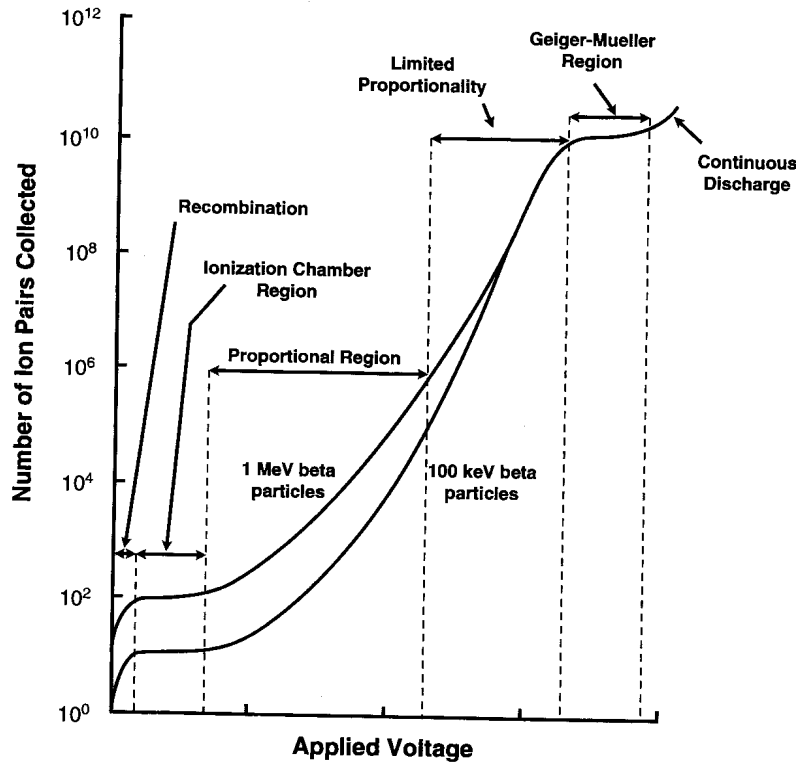


FIGURE 20-5. Amount of electrical charge collected after a single interaction as a function of the electrical potential difference (voltage) applied between the two electrodes of a gas-filled detector. The lower curve shows the charge collected when a 100-keV electron interacts, and the upper curve shows the result from a 1-MeV electron.

not significantly increase the current. Ionization chambers are operated in this region.

Beyond the ionization region, the collected current again increases as the applied voltage is raised. In this region, called the *proportional region*, electrons approaching the anode are accelerated to such high kinetic energies that they cause additional ionization. This phenomenon, called *gas multiplication*, amplifies the collected current; the amount of amplification increases as the applied voltage is raised.

At any voltage through the ionization chamber region and the proportional region, the amount of electrical charge collected from each interaction is *proportional* to the amount of energy deposited in the gas of the detector by the interaction. For example, the amount of charge collected after an interaction depositing 100 keV is one-tenth of that collected from an interaction depositing 1 MeV.

Beyond the proportional region is a region in which the amount of charge collected from each event is the same, regardless of the amount of energy deposited by the interaction. In this region, called the *Geiger-Mueller region* (GM region), the gas multiplication spreads the entire length of the anode. The size of a pulse in the GM region tells us nothing about the energy deposited in the detector by the interaction causing the pulse. Gas-filled detectors cannot be operated at voltages beyond the GM region because they continuously discharge.

Ionization Chambers (Ion Chambers)

Because gas multiplication does not occur at the relatively low voltages applied to ionization chambers, the amount of electrical charge collected from a single interaction is very small and would require a huge amplification to be detected. For this reason, ionization chambers are seldom used in pulse mode. The advantage to operating them in current mode is the freedom from dead-time effects, even in very intense radiation fields. In addition, as shown in Fig. 20-5, the voltage applied to an ion chamber can vary significantly without appreciably changing the amount of charge collected.

Almost any gas can be used to fill the chamber. If the gas is air and the walls of the chamber are of a material whose effective atomic number is similar to air, the amount of current produced is proportional to the *exposure rate* (exposure is the amount of electrical charge produced per mass of air). Air-filled ion chambers are used in portable survey meters and can accurately indicate exposure rates from less than 1 mR/hr to hundreds of roentgens per hour (Fig. 20-6). Air-filled ion chambers are also used for performing quality-assurance testing of diagnostic and therapeutic x-ray machines, and they are the detectors in most x-ray machine phototimers.

Gas-filled detectors tend to have low intrinsic efficiencies for detecting x-rays and gamma rays because of the low densities of gases and the low atomic numbers of most common gases. The sensitivity of ion chambers to x-rays and gamma rays can be enhanced by filling them with a gas that has a high atomic number, such as argon ($Z = 18$) or xenon ($Z = 54$), and pressurizing the gas to increase its density. Well-type ion chambers called dose calibrators are used in nuclear medicine to assay the activities of dosages of radiopharmaceuticals to be administered to patients; many are filled with pressurized argon. Xenon-filled pressurized ion chambers are used as detectors in some CT machines.

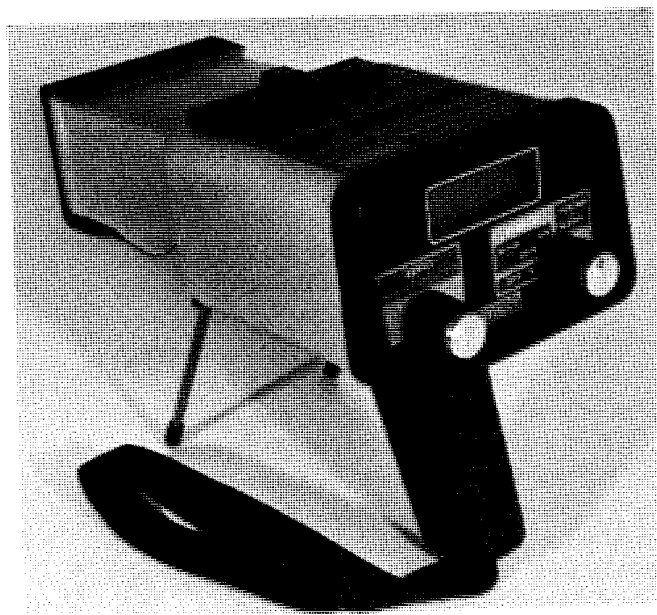


FIGURE 20-6. Portable air-filled ionization chamber survey meter. This particular instrument measures exposure rates ranging from about 0.1 to 20 R/hr. (Photograph courtesy Invision Radiation Measurements, Solon, Ohio.)

Proportional Counters

Unlike ion chambers, which can function with almost any gas, a proportional counter must contain a gas with specific properties, discussed in more advanced texts. Because gas amplification can produce a charge-per-interaction hundreds or thousands of times larger than that produced by an ion chamber, proportional counters can be operated in pulse mode as counters or spectrometers. They are commonly used in standards laboratories, in health physics laboratories, and for physics research. They are seldom used in medical centers.

Multiwire proportional counters, which indicate the position of an interaction in the detector, have been studied for use in nuclear medicine imaging devices. They have not achieved acceptance because of their low efficiencies for detecting x-rays and gamma rays from the radionuclides commonly used in nuclear medicine.

Geiger-Mueller Counters

GM counters must contain gases with specific properties, discussed in more advanced texts. Because gas amplification produces billions of ion pairs after an interaction, the signal from a GM detector requires little additional amplification. For this reason, GM detectors are often used for inexpensive survey meters. Flat, thin-window GM detectors, called “pancake” type detectors, are very useful for finding radioactive contamination (Fig. 20-7). The thin window allows most beta particles and conversion electrons to reach the gas and be detected.

GM detectors have a high efficiency for detecting charged particles. Almost every particle reaching the interior of the detector is counted. Very weak charged particles, such as the beta particles emitted by tritium (^3H , $E_{\text{max}} = 18 \text{ keV}$), which

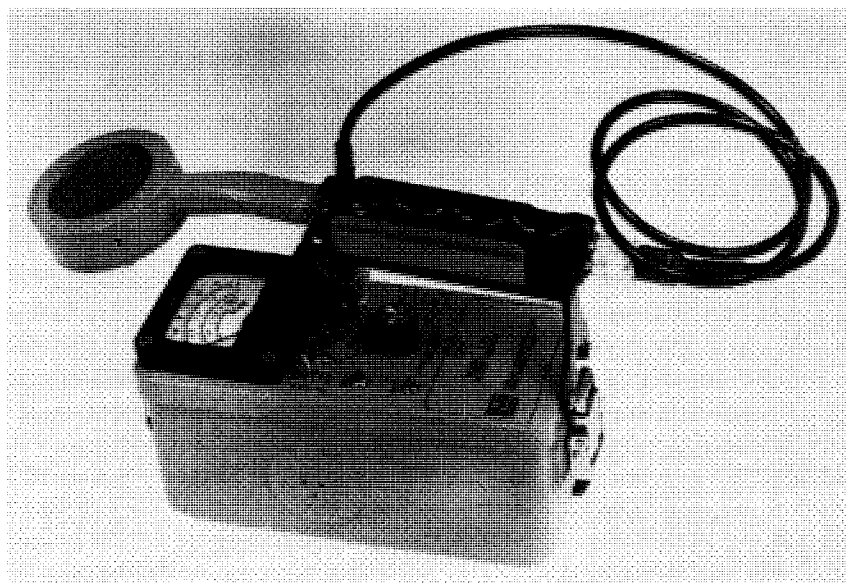


FIGURE 20-7. Portable Geiger-Mueller survey meter with a thin-window “pancake” probe. (Photograph courtesy Ludlum Measurements, Inc., Sweetwater, Texas)

is extensively used in biomedical research, do not penetrate the window; therefore, contamination by ^3H cannot be detected with a GM survey meter.

In general, GM survey meters are very inefficient detectors of x-rays and gamma rays. Those that are detected usually interact with the wall of the detector, with the resultant electrons scattered into the gas inside the detector.

The size of the voltage pulse from a GM tube is independent of the energy deposited in the detector by the interaction causing the pulse: an interaction that deposits 1 keV causes a voltage pulse of the same size as one caused by an interaction that deposits 1 MeV. Therefore, GM detectors cannot be used as spectrometers or dose-rate meters. Many portable GM survey meters display measurements in units of milliroentgens per hour. However, the GM counter cannot truly measure exposure rates, and so its reading must be considered only a crude approximation. If a GM survey meter is calibrated to indicate exposure rate for 662-keV gamma rays (commonly used for calibrations), it may over-respond by as much as a factor of 5 for photons of lower energies, such as 80 keV. If an accurate measurement of exposure rate is required, an air-filled ionization chamber survey meter should be used.

This over-response of a GM tube to low-energy x-rays and gamma rays can be partially corrected by placing a thin layer of a material with a higher atomic number (e.g., tin) around the detector. The increasing attenuation coefficient of the material (due to the photoelectric effect) with decreasing photon energy significantly flattens the energy response of the detector. Such GM tubes are called *energy-compensated detectors*. The disadvantage of an energy-compensated detector is that its sensitivity to lower-energy photons is substantially reduced and its energy threshold, below which photons cannot be detected at all, is increased. Energy compensated GM tubes often have windows that can be opened so that high-energy beta particles and low-energy photons can be detected.

GM detectors suffer from extremely long dead times, ranging from tens to hundreds of microseconds. For this reason, GM counters are seldom used when accurate measurements are required of count rates greater than a few hundred counts per second. A portable GM survey meter may become paralyzed in a very high radiation field and yield a reading of zero. Ionization chamber instruments should always be used to measure high radiation fields.

20.3 SCINTILLATION DETECTORS

Basic Principles

Scintillators are materials that emit visible or ultraviolet light after the interaction of ionizing radiation with the material. Scintillators are the oldest type of radiation detector; Roentgen discovered x-radiation and the fact that x-rays induce scintillation in barium platinocyanide in the same fortuitous experiment. Scintillators are used in conventional film-screen radiography, many digital radiographic image receptors, fluoroscopy, scintillation cameras, most CT scanners, and PET scanners.

Although the light emitted from a single interaction can be seen if the viewer's eyes are dark adapted, most scintillation detectors incorporate a means of signal amplification. In conventional film-screen radiography, photographic film is used to amplify and record the signal. In other applications, electronic devices such as photomultiplier tubes (PMTs), photodiodes, or image-intensifier tubes convert the light into an electrical signal. PMTs and image-intensifier tubes amplify the signal

as well. However, most photodiodes do not provide amplification; if amplification of the signal is required, it must be provided by an electronic amplifier. A *scintillation detector* consists of a scintillator and a device, such as a PMT, that converts the light into an electrical signal.

When ionizing radiation interacts with a scintillator, electrons are raised to an excited energy level. Ultimately, these electrons fall back to a lower energy state, with the emission of visible or ultraviolet light. Most scintillators have more than one mode for the emission of visible light, and each mode has its characteristic decay constant. *Luminescence* is the emission of light after excitation. *Fluorescence* is the prompt emission of light, whereas *phosphorescence* (also called *afterglow*) is the delayed emission of light. When scintillation detectors are operated in current mode, the prompt signal from an interaction cannot be separated from the phosphorescence caused by previous interactions. When a scintillation detector is operated in pulse mode, afterglow is less important because electronic circuits can separate the rapidly rising and falling components of the prompt signal from the slowly decaying delayed signal resulting from previous interactions.

It is useful, before discussing actual scintillation materials, to consider properties that are desirable in a scintillator.

1. The *conversion efficiency*, the fraction of deposited energy that is converted into light, should be high. (Conversion efficiency should not be confused with detection efficiency.)
2. For many applications, the decay times of excited states should be short. (Light is emitted promptly after an interaction.)
3. The material should be transparent to its own emissions. (Most emitted light escapes reabsorption.)
4. The frequency spectrum (color) of emitted light should match the spectral sensitivity of the light receptor (PMT, photodiode, or film).
5. If used for x-ray and gamma-ray detection, the attenuation coefficient (μ) should be large, so that detectors made of the scintillator have high detection efficiencies. (Materials with large atomic numbers and high densities have large attenuation coefficients.)
6. The material should be rugged, unaffected by moisture, and inexpensive to manufacture.

In all scintillators, the amount of light emitted after an interaction increases with the energy deposited by the interaction. Therefore, scintillators may be operated in pulse mode as spectrometers. When a scintillator is used for spectroscopy, its energy resolution (ability to distinguish between interactions depositing different energies) is primarily determined by its conversion efficiency. A high conversion efficiency produces superior energy resolution.

There are several categories of materials that scintillate. Many organic compounds exhibit scintillation. In these materials, the scintillation is a property of the molecular structure. Solid organic scintillators are used for timing experiments in particle physics because of their extremely prompt light emission. Organic scintillators include the liquid scintillation fluids that are used extensively in biomedical research. Samples containing radioactive tracers such as H-3, C-14, and P-32 are mixed in vials with liquid scintillators, and the number of light flashes are detected and counted with the use of PMTs and associated electronic circuits. Organic scintillators are not used for medical imaging because the low atomic numbers of their

constituent elements and their low densities make them poor x-ray and gamma-ray detectors. When photons in the diagnostic energy range do interact with organic scintillators, it is primarily by Compton scattering.

There are also many inorganic crystalline materials that exhibit scintillation. In these materials, the scintillation is a property of the crystalline structure: if the crystal is dissolved, the scintillation ceases. Many of these materials have much larger average atomic numbers and higher densities than organic scintillators and therefore are excellent photon detectors. They are widely used for radiation measurements and imaging in radiology.

Most inorganic scintillation crystals are deliberately grown with trace amounts of impurity elements called *activators*. The atoms of these activators form preferred sites in the crystals for the excited electrons to return to the ground state. The activators modify the frequency (color) of the emitted light, the promptness of the light emission, and the proportion of the emitted light that escapes reabsorption in the crystal.

Inorganic Crystalline Scintillators in Radiology

No one scintillation material is best for all applications in radiology. Sodium iodide activated with thallium [NaI(Tl)] is used for most nuclear medicine applications. It is coupled to PMTs and operated in pulse mode in scintillation cameras, thyroid probes, and gamma well counters. Its high content of iodine ($Z = 53$) and high density provide a high photoelectric absorption probability for x-rays and gamma rays emitted by common nuclear medicine radiopharmaceuticals (70 to 365 keV). It has a very high conversion efficiency; approximately 13% of deposited energy is converted into light. Because a light photon has an energy of about 3 eV, approximately one light photon is emitted for every 23 eV absorbed by the crystal. This high conversion efficiency gives it one of the best energy resolutions of any scintillator. It emits light very promptly (decay constant, 230 nsec), permitting it to be used in pulse mode at interaction rates greater than 100,000/sec. Very large crystals can be manufactured; for example, the rectangular crystals of one modern scintillation camera are 59 cm (23 inches) long, 44.5 cm (17.5 inches) wide, and 0.95 cm thick. Unfortunately, NaI(Tl) crystals are fragile; they crack easily if struck or subjected to rapid temperature change. Also, they are hygroscopic (i.e., they absorb water from the atmosphere) and therefore must be hermetically sealed.

Crystals of bismuth germanate ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$, often abbreviated "BGO") are coupled to PMTs and used in pulse mode as detectors in most PET scanners. The high atomic number of bismuth ($Z = 83$) and the high density of the crystal yield a high intrinsic efficiency for the 511-keV positron annihilation photons. The primary component of the light emission is sufficiently prompt (decay constant, 300 nsec) to permit annihilation coincidence detection, which is the basis for PET (see Chapter 22). NaI(Tl) is used in some less-expensive PET scanners, and lutetium oxyorthosilicate and gadolinium oxyorthosilicate, both activated with cerium, are used in some newer PET scanners.

Calcium tungstate (CaWO_4) was used for many years in intensifying screens in film-screen radiography. It has been largely replaced by rare-earth phosphors, such as gadolinium oxysulfide activated with terbium and yttrium tantalate. The intensifying screen is an application of scintillators that does not require prompt light emission, because the film usually remains in contact with the screen for at least several seconds. Cesium iodide activated with thallium is used as the phosphor layer of

TABLE 20-1. INORGANIC SCINTILLATORS USED IN MEDICAL IMAGING

Material	Atomic numbers	Density (g/cm ³)	Wavelength of maximal emission (nm)	Conversion efficiency ^a (%)	Decay constant (μs)	Afterglow after 3 msec (%)	Uses
NaI(Tl)	11, 53	3.67	415	100	0.23	0.3–5	Scintillation cameras
Bi ₄ Ge ₃ O ₁₂	83, 32, 8	7.13	480	12–14	0.3	0.1	Position emission tomography scanners
CsI(Na)	55, 53	4.51	420	85	0.63	0.5–5	Input phosphor of image intensifier tubes
CsI(Tl)	55, 53	4.51	565	45 ^b	1.0	0.5–5	Indirect detection thin-film transistor radiographic image receptors
ZnCdS(Ag)	30, 48, 16	—	—	—	—	—	Output phosphor of image intensifier tubes
CdWO ₄	48, 74, 8	7.90	540	40	5	0.1	Computed tomographic scanners
CaWO ₄	20, 74, 8	6.12	—	14–18	0.9–20	—	Radiographic screens
Gd ₂ O ₂ S(Tb)	64, 8, 16	7.34	—	—	560	—	Radiographic screens

^aRelative to NaI(Tl), using a photomultiplier tube to measure light.

^bThe light emitted by CsI(Tl) does not match the spectral sensitivity of photomultiplier tubes very well; its conversion efficiency is much larger, if measured with a photodiode.

Source: Data courtesy Bicron, Saint-Gobain/Norton Industrial Ceramics Corporation.

many indirect-detection thin-film transistor radiographic image receptors, described in Chapter 11. Cesium iodide activated with sodium is used as the input phosphor and zinc cadmium sulfide activated with silver is used as the output phosphor of image-intensifier tubes.

Scintillators coupled to photodiodes are used as the detectors in most CT scanners. The extremely high x-ray flux experienced by the detectors necessitates current mode operation to avoid dead-time effects. With the rotational speed of CT scanners approaching half a second, the scintillators used in CT must have very little afterglow. Cadmium tungstate and gadolinium ceramics are scintillators commonly used in CT. Table 20-1 lists the properties of several inorganic crystalline scintillators of importance in radiology and nuclear medicine.

Conversion of Light into an Electrical Signal

Photomultiplier Tubes

PMTs perform two functions—conversion of ultraviolet and visible light photons into an electrical signal and signal amplification, on the order of millions to billions. As shown in Fig. 20-8, a PMT consists of an evacuated glass tube containing a *photocathode*, typically 10 to 12 electrodes called *dynodes*, and an *anode*. The photo-

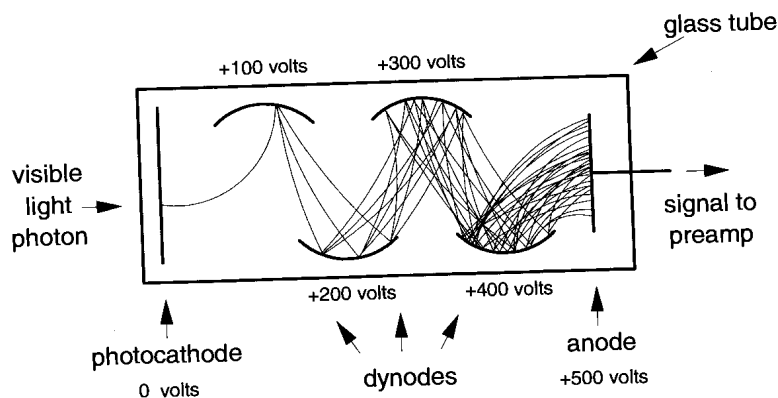


FIGURE 20-8. Photomultiplier tube. Actual photomultiplier tubes typically have 10 to 12 dynodes.

cathode is a very thin electrode, located just inside the glass entrance window of the PMT, which emits electrons when struck by visible light. (Approximately one electron is emitted from the photocathode for every five light photons incident upon it.) A high-voltage power supply provides a voltage of approximately 1,000 V, and a series of resistors divides the voltage into equal increments. The first dynode is given a voltage of about +100 V with respect to the photocathode; successive dynodes have voltages that increase by approximately 100 V per dynode. The electrons emitted by the photocathode are attracted to the first dynode and are accelerated to kinetic energies equal to the potential difference between the photocathode and the first dynode. (If the potential difference is 100 V, the kinetic energy of each electron is 100 eV.) When these electrons strike the first dynode, about five electrons are ejected from the dynode for each electron hitting it. These electrons are then attracted to the second dynode, reaching kinetic energies equal to the potential difference between the first and second dynodes, and causing about five electrons to be ejected from the second dynode for each electron hitting it. This process continues down the chain of dynodes, with the number of electrons being multiplied by a factor of 5 at each stage. The total amplification of the PMT is the product of the individual multiplications at each dynode. If a PMT has ten dynodes and the amplification at each stage is 5, the total amplification will be:

$$5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 = 5^{10} \approx 10,000,000$$

The amplification can be adjusted by changing the voltage applied to the PMT.

When a scintillator is coupled to a PMT, an optical coupling material is placed between the two components to minimize reflection losses. The scintillator is usually surrounded on all other sides by a highly reflective material, often magnesium oxide powder.

Photodiodes

Photodiodes are semiconductor diodes that convert light into an electrical signal. (The principles of operation of semiconductor diodes are discussed later in this chapter.) In use, photodiodes are reverse biased. *Reverse bias* means that the voltage is applied with the polarity such that essentially no electrical current flows. When the photodiode is exposed to light, an electrical current is generated that is propor-

tional to the intensity of the light. Photodiodes are sometimes used with scintillators instead of PMTs. Photodiodes produce more electrical noise than PMTs do, but they are smaller and less expensive. Most photodiodes, unlike PMTs, do not amplify the signal. Photodiodes coupled to CdWO_4 or other scintillators are used in current mode in many CT scanners. Photodiodes are also essential components of indirect-detection thin-film transistor radiographic image receptors, which use scintillators to convert x-ray energy into light.

Scintillators with Trapping of Excited Electrons

Thermoluminescent Dosimeters

In most applications of scintillators, the prompt emission of light after an interaction is desirable. However, there are inorganic scintillators, called *thermoluminescent dosimeters* (TLDs), in which electrons become trapped in excited states after interactions with ionizing radiation. If the scintillator is later heated, the electrons then fall to their ground state with the emission of light. This property makes these materials useful as dosimeters.

In use, a TLD is exposed to ionizing radiation. It is later “read” by heating it in front of a PMT. The amount of light emitted by the TLD is proportional to the amount of energy absorbed by the TLD. After the TLD has been read, it may be baked in an oven and reused.

Lithium fluoride (LiF) is one of the most useful TLD materials. It exhibits almost negligible release of trapped electrons at room temperature, so there is little loss of information with time from exposure to the reading of the TLD. The effective atomic number of LiF is close to that of tissue, so the amount of light emission is almost proportional to the tissue dose over the range of x-ray and gamma-ray energies. It is often used instead of photographic film for personnel dosimetry.

Photostimulable Phosphors

Photostimulable phosphors (PSPs), like TLDs, are scintillators in which a fraction of the excited electrons become trapped. PSP plates are used in radiography as image receptors, instead of film-screen cassettes. Although the trapped electrons could be released by heating, a laser is used to scan the plate and release them. The electrons then fall to the ground state, with the emission of light. Barium fluorohalide activated with europium is commonly used for PSP imaging plates. In this material, the wavelength that is most efficient in stimulating luminescence is in the red portion of the spectrum, whereas the stimulated luminescence itself is in the blue-violet portion of the spectrum. The stimulated emissions are converted into an electrical signal by PMTs. After the plate is read by the laser, it may be exposed to light to release the remaining trapped electrons and reused. The use of PSPs in radiography is discussed further in Chapter 11.

20.4 SEMICONDUCTOR DETECTORS

Semiconductors are crystalline materials whose electrical conductivities are less than those of metals but more than those of crystalline insulators. Silicon and germanium are common semiconductor materials.

In crystalline materials, electrons exist in energy bands, separated by forbidden gaps. In metals (e.g., copper), the least tightly bound electrons exist in a partially occupied band, called the conduction band. The conduction band electrons are mobile, providing high electrical conductivity. In an insulator or a semiconductor, the valence electrons exist in a filled valence band. In semiconductors, these valence band electrons participate in covalent bonds and so are immobile. The next higher energy band, the conduction band, is empty of electrons. However, if an electron is placed in the conduction band, it is mobile, as are the upper band electrons in metals. The difference between insulators and semiconductors is the magnitude of the energy gap between the valence and conduction bands. In insulators the band gap is greater than 5 eV, whereas in semiconductors it is about 1 eV or less (Fig. 20-9). In semiconductors, valence band electrons can be raised to the conduction band by ionizing radiation, visible or ultraviolet light, or thermal energy.

When a valence band electron is raised to the conduction band, it leaves behind a vacancy in the valence band. This vacancy is called a *hole*. Because a hole is the absence of an electron, it is considered to have a net positive charge, equal but opposite to that of an electron. When another valence band electron fills the hole, a hole is created at that electron's former location. Thus, holes behave as mobile positive charges in the valence band. The hole-electron pairs formed in a semiconductor material by ionizing radiation are analogous to the ion pairs formed in a gas by ionizing radiation.

A crystal of a semiconductor material can be used as a radiation detector. A voltage is placed between two terminals on opposite sides of the crystal. When ionizing radiation interacts with the detector, electrons in the crystal are raised to an excited state, permitting an electrical current to flow, similar to a gas-filled ionization chamber. Unfortunately, the radiation-induced current, unless it is very large, is masked by a larger current induced by the applied voltage.

To reduce the magnitude of the voltage-induced current so that the signal from radiation interactions can be detected, the semiconductor crystal is "doped" with a trace amount of impurities so that it acts as a diode (see earlier discussion of types of detectors). The impurity atoms fill sites in the crystal lattice that would otherwise be occupied by atoms of the semiconductor material. If atoms of the impurity material have more valence electrons than those of the semiconductor material, the

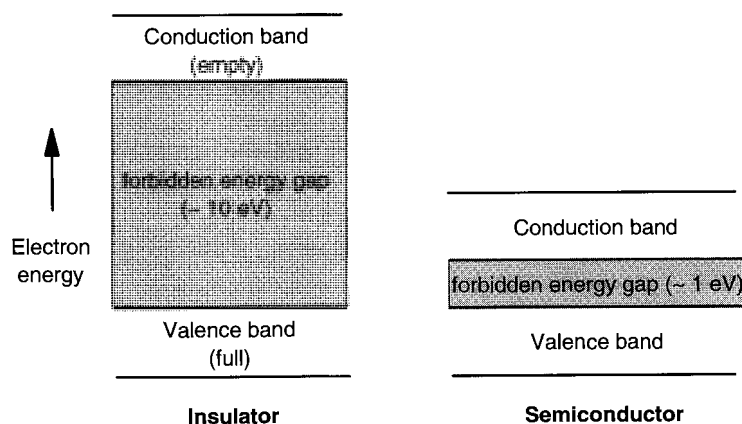


FIGURE 20-9. Energy band structure of a crystalline insulator and a semiconductor material.

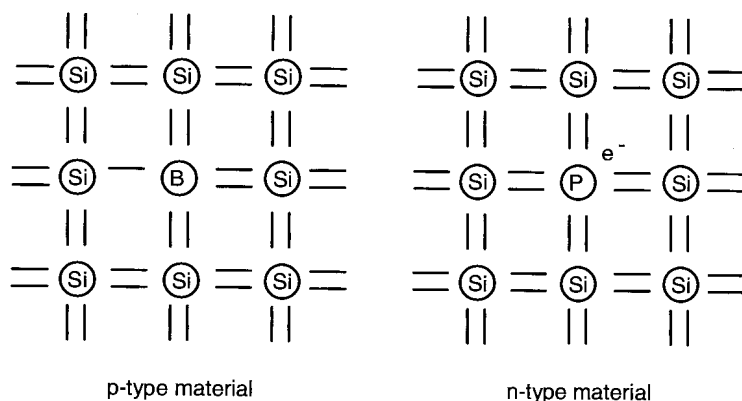


FIGURE 20-10. P-type and n-type impurities in a crystal of a semiconductor material. N-type impurities provide mobile electrons in the conduction band, whereas p-type impurities provide acceptor sites in the valence band. When filled by electrons, these acceptor sites create holes that act as mobile positive charges.

impurity atoms provide mobile electrons in the conduction band. A semiconductor material containing an electron-donor impurity is called an *n-type material* (Fig. 20-10). N-type material has mobile electrons in the conduction band. On the other hand, an impurity with fewer valence electrons than the semiconductor material provides sites in the valence band that can accept electrons. When a valence band electron fills one of these sites, it creates a hole at its former location. Semiconductor material doped with a hole-forming impurity is called *p-type material*. P-type material has mobile holes in the valence band.

A semiconductor diode consists of a crystal of semiconductor material with a region of n-type material that forms a junction with a region of p-type material (Fig. 20-11). If an external voltage is applied with the positive polarity on the p-type side

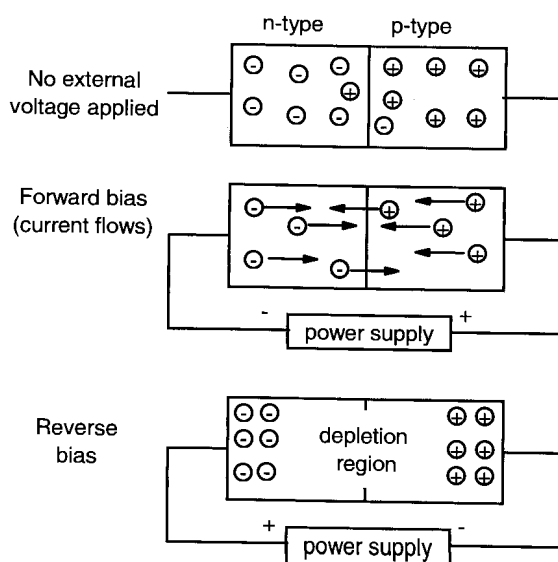


FIGURE 20-11. Semiconductor diode. When no bias is applied, a few holes migrate into the n-type material and a few conduction-band electrons migrate into the p-type material. With forward bias, the external voltage is applied with the positive polarity on the p-side of the junction and negative polarity on the n-side, causing the charge carriers to be swept into the junction and a large current to flow. With negative bias, the charge carriers are drawn away from the junction, creating a region depleted of charge carriers that acts as a solid-state ion chamber.

of the diode and the negative polarity on the n-type side, the holes in the p-type material and the mobile conduction-band electrons of the n-type material are drawn to the junction. There, the mobile electrons fall into the valence band to fill holes. Applying an external voltage in this manner is referred to as *forward bias*. Forward bias permits a current to flow with little resistance.

On the other hand, if an external voltage is applied with the opposite polarity—that is, the with the negative polarity on the p-type side of the diode and the positive polarity on the n-type side—the holes in the p-type material and the mobile conduction-band electrons of the n-type material are drawn away from the junction. Applying the external voltage in this polarity is referred to as *reverse bias*. Reverse bias draws the charge carriers away from the n-p junction, forming a region depleted of current carriers. Very little electrical current flows when a diode is reverse-biased.

A reverse-biased semiconductor diode can be used to detect visible and ultraviolet light or ionizing radiation. The photons of light or ionization and excitation produced by ionizing radiation can excite lower-energy electrons in the depletion region of the diode to higher-energy bands, producing hole-electron pairs. The electrical field in the depletion region sweeps the holes toward the p-type side and the conduction band electrons toward the n-type side, causing a momentary pulse of current to flow after the interaction.

Photodiodes are semiconductor diodes that convert light into an electrical current. As mentioned previously, they are used in conjunction with scintillators as detectors in CT scanners. Scintillation-based thin-film transistor radiographic image receptors, discussed in Chapter 11, incorporate a photodiode in each pixel.

Semiconductor detectors are semiconductor diodes designed for the detection of ionizing radiation. The amount of charge generated by an interaction is proportional to the energy deposited in the detector by the interaction; therefore, semiconductor detectors are spectrometers. Because thermal energy can also raise electrons to the conduction band, many types of semiconductor detectors used for x-ray and gamma-ray spectroscopy must be cooled with liquid nitrogen.

Semiconductor detectors are seldom used for medical imaging devices because of high expense, low quantum detection efficiency in comparison to scintillators such as NaI(Tl) (Z of iodine = 53, Z of germanium = 32, Z of silicon = 14), because they can be manufactured only in limited sizes, and because many such devices require cooling.

Efforts are being made to develop semiconductor detectors of higher atomic number than germanium that can be operated at room temperature. The leading candidate to date is cadmium zinc telluride (CZT). A small-field-of-view nuclear medicine camera using CZT detectors has been developed.

The energy resolution of germanium semiconductor detectors is greatly superior to that of NaI(Tl) scintillation detectors. Liquid nitrogen-cooled germanium detectors are widely used for the identification of individual gamma ray-emitting radionuclides in mixed radionuclide samples because of their superb energy resolution.

20.5 PULSE HEIGHT SPECTROSCOPY

Many radiation detectors, such as scintillation detectors, semiconductor detectors, and proportional counters, produce electrical pulses whose amplitudes are propor-

tional to the energies deposited in the detector by individual interactions. *Pulse height analyzers* (PHAs) are electronic systems that may be used with these detectors to perform pulse height spectroscopy and energy-selective counting. In energy-selective counting, only interactions that deposit energies within a certain energy range are counted. Energy-selective counting can be used to reduce the effects of background radiation, to reduce the effects of scatter, or to separate events caused by different radionuclides in a mixed radionuclide sample. Two types of PHAs are *single-channel analyzers* (SCAs) and *multichannel analyzers* (MCAs). MCAs determine spectra much more efficiently than do SCA systems, but they are more expensive. Pulse height discrimination circuits are incorporated in scintillation cameras and other nuclear medicine imaging devices to reduce the effects of scatter on the images.

Single-Channel Analyzer Systems

Function of a Single-Channel Analyzer System

Figure 20-12 depicts an SCA system. Although the system is shown with a NaI(Tl) crystal and PMT, it could be used with any pulse-mode spectrometer. The high-voltage power supply typically provides 800 to 1,200 volts to the PMT. The series of resistors divides the total voltage into increments that are applied to the dynodes and anode of the PMT. Raising the voltage increases the magnitude of the voltage pulses from the PMT.

The detector is often located some distance from the majority of the electronic components. The pulses from the PMT are usually routed to a preamplifier (pre-amp), which is connected to the PMT by as short a cable as possible. The function of the preamp is to amplify the voltage pulses further, so as to minimize distortion and attenuation of the signal during transmission to the remainder of the system. The pulses from the preamp are routed to the amplifier, which further amplifies the pulses and modifies their shapes. The gains of most amplifiers are adjustable.

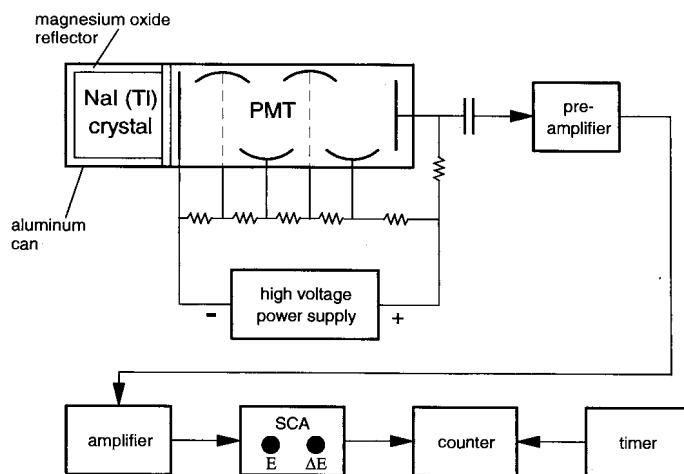


FIGURE 20-12. Single-channel analyzer system with NaI(Tl) detector and photomultiplier tube.

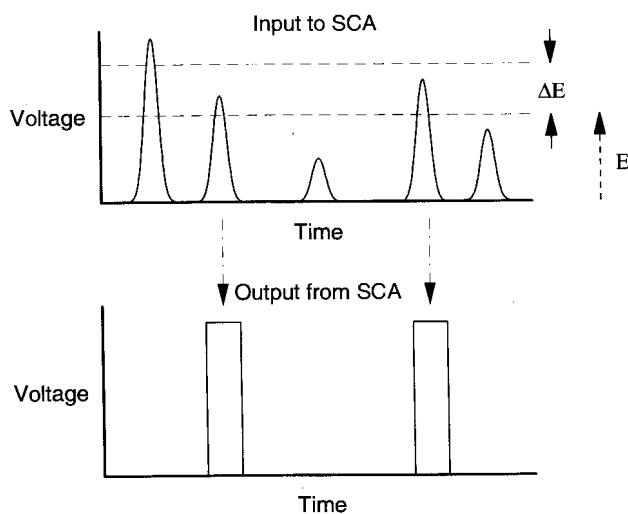


FIGURE 20-13. Function of a single-channel analyzer. Energy discrimination occurs by rejection of pulses above or below the energy window set by operator.

The pulses from the amplifier then proceed to the SCA. The user is allowed to set two voltage levels, a lower level and an upper level. If a voltage pulse whose amplitude is less than the lower level or greater than the upper level is received from the amplifier, the SCA does nothing. If a voltage pulse whose amplitude is greater than the lower level but less than the upper level is received from the amplifier, the SCA produces a single logic pulse. (A logic pulse is a voltage pulse of fixed amplitude and duration.) Figure 20-13 illustrates the operation of a SCA. The counter counts the logic pulses from the SCA for a time interval set by the timer.

Many SCAs permit the user to select the mode by which the two knobs set the lower and upper levels. In one mode, usually called *LL/UL mode*, one knob directly sets the lower level and the other sets the upper level. In another mode, called *window mode*, one knob (often labeled *E*) sets the midpoint of the range of acceptable pulse heights and the other knob (often labeled ΔE or *window*) sets the range of voltages around this value. In this mode, the lower-level voltage is $E - \Delta E/2$ and the upper level voltage is $E + \Delta E/2$. (In some SCAs, the range of acceptable pulse heights is from E to $E + \Delta E$.) Window mode is convenient for plotting a spectrum.

Plotting a Spectrum Using a Single-Channel Analyzer

To obtain the pulse height spectrum of a sample of radioactive material using an SCA system, the SCA is placed in window mode, the *E* setting is set to zero, and a small window (ΔE) setting is selected. A series of counts is taken for a fixed length of time per count, with the *E* setting increased before each count but without changing the window setting. Each count is plotted on graph paper as a function of baseline (*E*) setting.

Energy Calibration of a Single-Channel Analyzer System

On most SCAs, each of the two knobs permits values from 0 to 1,000 to be selected. By adjusting the amplification of the pulses reaching the SCA—either by

changing the voltage produced by the high-voltage power supply or by changing the amplifier gain—the system can be calibrated so that these knob settings directly indicate keV.

A cesium 137 (Cs-137) source, which emits 662-keV gamma rays, is usually used. A narrow window is set, centered about a setting of 662. For example, the SCA may be placed into LL/UL mode with lower-level value of 655 and an upper-level value of 669 selected. Then the voltage produced by the high-voltage power supply is increased in steps, with a count taken after each step. The counts first increase and then decrease. When the voltage that produces the largest count is selected, the two knobs on the SCA directly indicate keV. This procedure is called *peaking* the SCA system.

Multichannel Analyzer Systems

An MCA system permits an energy spectrum to be automatically acquired much more quickly and easily than does a SCA system. Figure 20-14 shows an MCA; Fig. 20-15 is a diagram of a counting system using an MCA. The detector, high-voltage power supply, preamp, and amplifier are the same as was described for SCA systems. The MCA consists of an analog-to-digital converter, a memory containing many storage locations called *channels*, control circuitry, a timer, and a cathode ray tube display. The memory of a MCA typically ranges from 256 to 8,192 channels, each of which can store a single integer. When the acquisition of a spectrum is begun, all of the channels are set to zero. When each voltage pulse from the amplifier is received, it is converted into a binary digital signal, the value of which is proportional to the amplitude of the analog voltage pulse. (Analog-to-digital converters are discussed in detail in Chapter 4.) This digital signal designates a particular channel in the MCA's memory. The number stored in that channel is then incremented by 1. As many pulses are processed, a spectrum is generated in the memory of the MCA. Figure 20-16 illustrates the operation of an MCA. Today, most MCAs are interfaced to digital computers that store, process, and display the resultant spectra.



FIGURE 20-14. Multichannel analyzer. (Photograph courtesy Canberra Industries, Inc.)

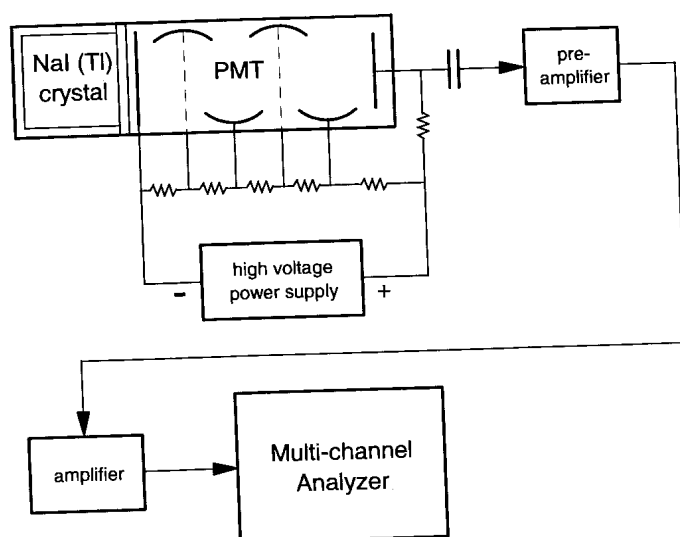


FIGURE 20-15. Multichannel analyzer system with NaI(Tl) detector.

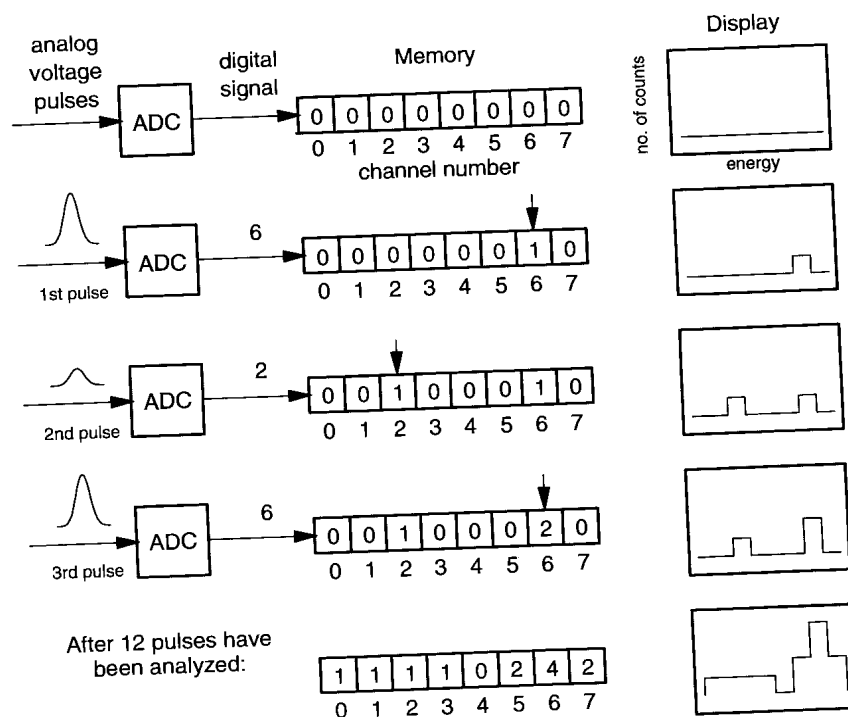


FIGURE 20-16. Acquisition of a spectrum by a multichannel analyzer (MCA). The digital signal produced by the analog-to-digital converter (ADC) is a set of binary pulses, as described in Chapter 4. After the analog pulses are digitized by the ADC, they are sorted into bins (channels) by height, forming an energy spectrum. Although this figure depicts an MCA with 8 channels, actual MCAs have as many as 8,192 channels.

X-Ray and Gamma-Ray Spectroscopy with Sodium Iodide Detectors

X-ray and gamma-ray spectroscopy is best performed with semiconductor detectors because of their superior energy resolution. However, high detection efficiency is more important than ultra-high energy resolution for most nuclear medicine applications, so most spectroscopy systems in nuclear medicine use NaI(Tl) crystals coupled to PMTs.

Interactions of Photons with a Spectrometer

There are a number of mechanisms by which an x-ray or gamma ray can deposit energy in the detector, several of which deposit only a fraction of the incident photon energy. As shown in Fig. 20-17, an incident photon can deposit its full energy by a photoelectric interaction (A) or by one or more Compton scatters followed by a photoelectric interaction (B). However, a photon will deposit only a fraction of its energy if it interacts by Compton scattering and the scattered photon escapes the detector (C). In that case, the energy deposited depends on the scattering angle, with larger angle scatters depositing larger energies. Even if the incident photon interacts by the photoelectric effect, less than its total energy will be deposited if the inner-shell electron vacancy created by the interaction results in the emission of a characteristic x-ray that escapes the detector (D).

Most detectors are shielded to reduce the effects of natural background radiation and nearby radiation sources. Figure 20-17 shows two ways by which a x-ray or gamma-ray interaction in the shield of the detector can deposit energy in the detector. The photon may Compton scatter in the shield, with the scattered photon striking the detector (E), or a characteristic x-ray from the shield may interact with the detector (F).

Most interactions of x-rays and gamma rays with an NaI(Tl) detector are with iodine atoms, because iodine has a much larger atomic number than sodium does. Although thallium has an even larger atomic number, it is only a trace impurity.

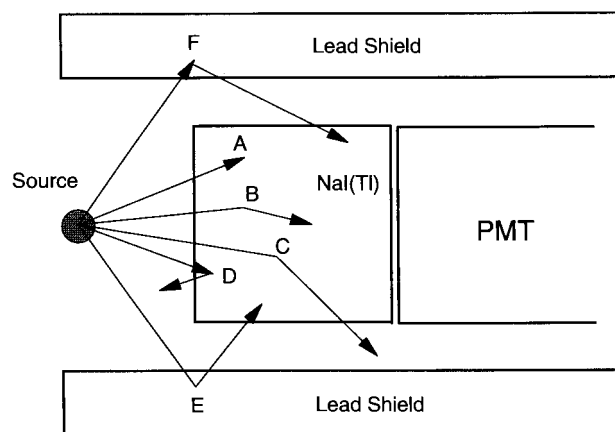


FIGURE 20-17. Interactions of x-rays and gamma rays with an NaI(Tl) detector. See text for description.

Spectrum of Cesium-137

The spectrum of Cs-137 is often used to introduce pulse height spectroscopy because of the simple decay scheme of this radionuclide. As shown at the top of Fig. 20-18, Cs-137 decays by beta particle emission to barium-137m (Ba-137m), leaving the Ba-137m nucleus in an excited state. The Ba-137m nucleus attains its ground state by the emission of a 662-keV gamma ray 90% of the time. In 10% of the decays, a conversion electron is emitted instead of a gamma ray. The conversion electron is usually followed by a ~32-keV *K*-shell characteristic x-ray as an outer-shell electron fills the inner-shell vacancy.

In the left in Fig. 20-18 is the actual energy spectrum of Cs-137, and on the right is its pulse height spectrum obtained with the use of an NaI(Tl) detector. There are two reasons for the differences between the spectra. First, there are a number mechanisms by which an x-ray or gamma ray can deposit energy in the detector, several of which deposit only a fraction of the incident photon energy. Second, there are random variations in the processes by which the energy deposited in the detector is converted into an electrical signal. In the case of an NaI(Tl) crystal coupled to a PMT, there are random variations in the fraction of deposited energy converted into light, the fraction of the light that reaches the photocathode of the PMT, and the number of electrons ejected from the back of the photocathode per unit energy deposited by the light. These factors cause random variations in the size of the voltage pulses produced by the detector, even when the incident x-rays or

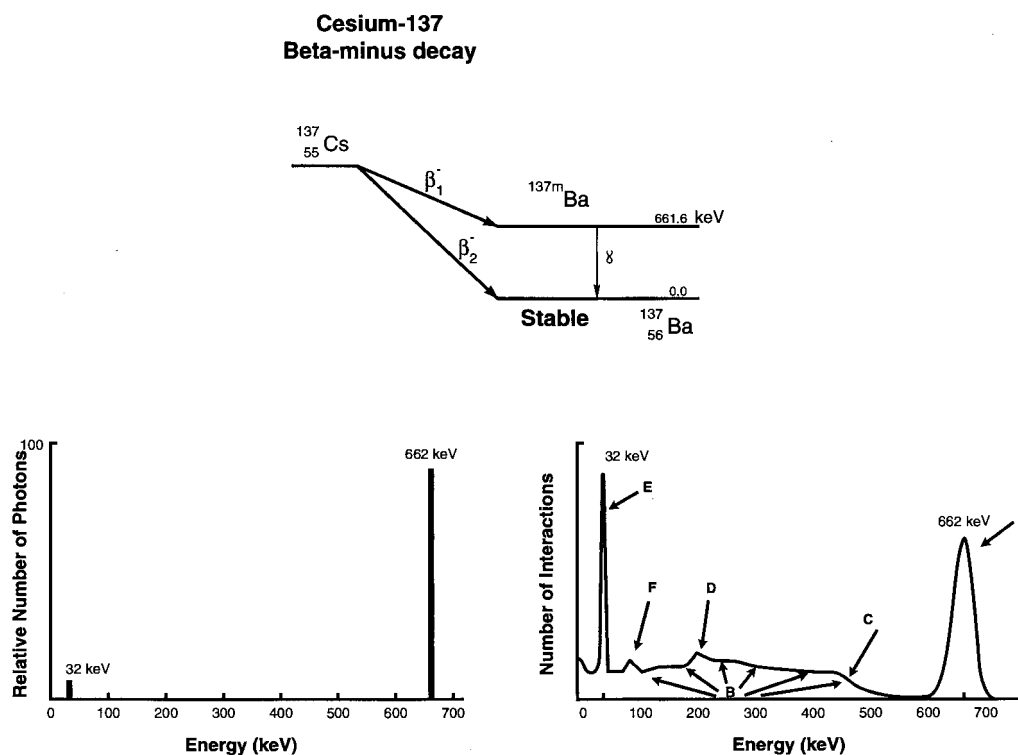


FIGURE 20-18. Decay scheme of cesium 137 (top), actual energy spectrum (left), and pulse height spectrum obtained using an NaI(Tl) scintillation detector (right). See text for description of pulse height spectrum.

gamma rays deposit exactly the same energy. The energy resolution of a spectrometer is a measure of the effect of these random variations on the resultant spectrum.

In the pulse height spectrum of Cs-137, on the right in Fig. 20-18, the photopeak (A) corresponds to interactions in which the energy of an incident 662-keV photon is entirely absorbed in the crystal. This may occur by a single photoelectric interaction or by one or more Compton interactions followed by a photoelectric interaction. The Compton continuum (B) is caused by 662-keV photons that scatter in the crystal, with the scattered photons escaping the crystal. Each portion of the continuum corresponds to a particular scattering angle. The Compton edge (C) is the upper limit of the Compton continuum. The backscatter peak (D) is caused by 662-keV photons that scatter from the shielding around the detector into the detector. The barium x-ray photopeak (E) is a second photopeak caused by the absorption of barium *K*-shell x-rays (31 to 37 keV), which are emitted after the emission of conversion electrons. Another photopeak (F) is caused by lead *K*-shell x-rays (72 to 88 keV) from the shield.

Spectrum of Technetium-99m

The decay scheme of technetium 99m (Tc-99) is shown at the top in Fig. 20-19. Tc-99 is an isomer of ^{99}Tc that decays by isomeric transition to its ground state,

TECHNETIUM 99m Isomeric Transition Decay

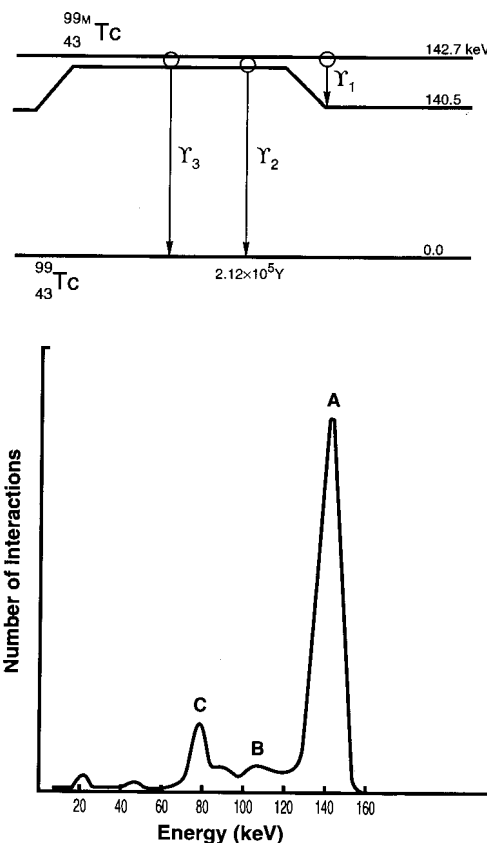


FIGURE 20-19. Decay scheme of technetium 99m (top) and its pulse height spectrum on an NaI(Tl) scintillation detector (bottom). See text for details.

with the emission of a 140.5-keV gamma ray. In 11% of the transitions, a conversion electron is emitted instead of a gamma ray.

The pulse height spectrum of Tc-99 is shown at the bottom of Fig. 20-19. The photopeak (A) is caused by the total absorption of the 140-keV gamma rays. The escape peak (B) is caused by 140-keV gamma rays that interact with the crystal by the photoelectric effect but with the resultant iodine *K*-shell x-rays (28 to 33 keV) escaping the crystal. There is also a photopeak (C) caused by the absorption of lead *K*-shell x-rays from the shield. The Compton continuum is quite small, unlike the continuum in the spectrum of Cs-137, because the photoelectric effect predominates in iodine at 140 keV.

Spectrum of Iodine-125

Iodine-125 (I-125) decays by electron capture followed by the emission of a 35.5-keV gamma ray (6.5% of decays) or a conversion electron. The electron capture usually leaves the daughter nucleus with a vacancy in the *K*-shell. The emission of a conversion electron usually also results in a *K*-shell vacancy. Each transformation of an I-125 atom therefore results in the emission, on the average, of 1.44 x-rays or gamma rays with energies between 27 and 36 keV.

Figure 20-20 shows two pulse height spectra from I-125. The spectrum on the left was acquired with the source located 7.5 cm from an NaI(Tl) detector, and the one on the right was collected with the source in an NaI(Tl) well detector. The spec-

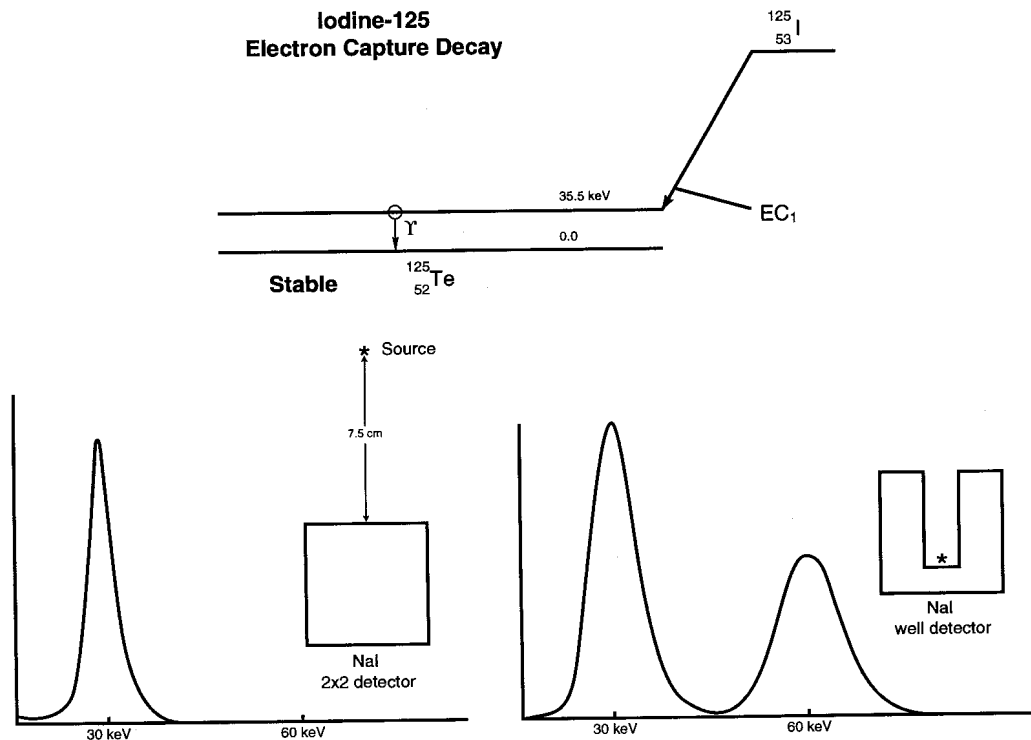


FIGURE 20-20. Spectrum of iodine 125 source located 7.5 cm from solid NaI(Tl) crystal (*left*) and in NaI(Tl) well counter (*right*).

trum on the left shows a large photopeak at about 30 keV, whereas the spectrum on the right shows a peak at about 30 keV and a smaller peak at about 60 keV. The 60-keV peak in the spectrum from the well detector is a *sum peak* caused by two photons simultaneously striking the detector. The sum peak is not apparent in the spectrum with the source 7.5 cm from the detector because the much lower detection efficiency renders unlikely the simultaneous interaction of two photons with the detector.

Performance Characteristics

Energy Resolution

The energy resolution of a spectrometer is a measure of its ability to differentiate between particles or photons of different energies. It can be determined by irradiating the detector with monoenergetic particles or photons and measuring the width of the resultant peak in the pulse height spectrum. Statistical effects in the detection process cause the amplitudes of the pulses from the detector to randomly vary about the mean pulse height, giving the peak a Gaussian shape. (These statistical effects are one reason why the pulse height spectrum produced by a spectrometer is not identical to the actual energy spectrum of the radiation.) A wider peak implies a poorer energy resolution. The width is usually measured at half the maximal height of the peak, as illustrated in Fig. 20-21. This is called the full width at half-maximum (FWHM). The FWHM is then divided by the pulse amplitude (not the peak height) corresponding to the maximum of the peak:

$$\text{Energy resolution} = \frac{\text{FWHM}}{\text{Pulse height at center of peak}} \times 100\% \quad [20-4]$$

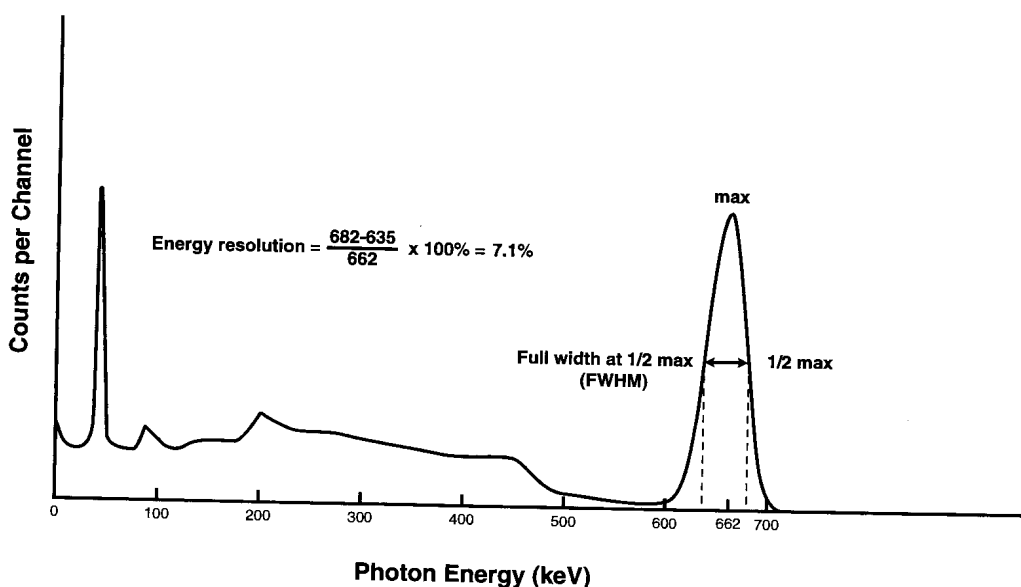


FIGURE 20-21. Energy resolution of a pulse height spectrometer. The spectrum shown is that of cesium 137, obtained by an NaI(Tl) scintillator coupled to a photomultiplier tube.

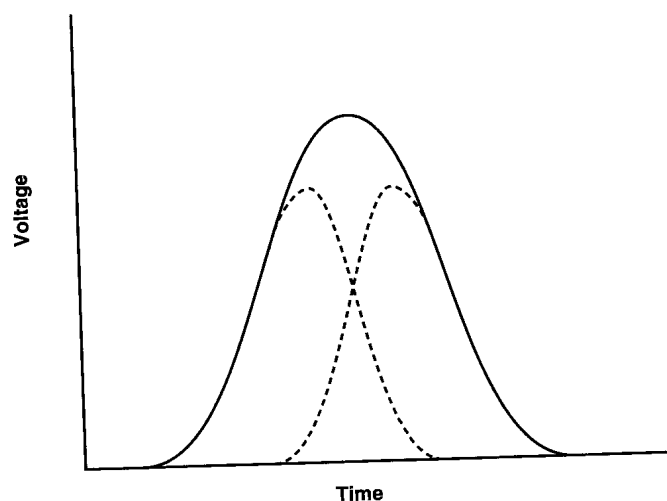


FIGURE 20-22. Pulse pileup. The dashed lines represent the signals produced by two individual interactions in the detector that occur at almost the same time. The solid line depicts the actual pulse produced by the detector. This single pulse is the sum of the signals from the two interactions.

For example, the energy resolution of a 5-cm-diameter and 5-cm-thick NaI(Tl) crystal, coupled to a PMT and exposed to the 662-keV gamma rays of Cs-137, is typically about 7% to 8%.

Count-Rate Effects in Spectrometers

In pulse height spectroscopy, count-rate effects are best understood as pulse pileup. Fig. 20-22 depicts the signal from a detector in which two interactions occur, separated by a very short time interval. The detector produces a single pulse, which is the sum of the individual signals from the two interactions, having a higher amplitude than the signal from either individual interaction. Because of this effect, operating a pulse height spectrometer at a high count rate causes loss of counts and misplacement of counts in the spectrum.

20.6 NONIMAGING DETECTOR APPLICATIONS

Sodium Iodide Thyroid Probe and Well Counter

Thyroid Probe

Most nuclear medicine departments have a thyroid probe for measuring the uptake of I-123 or I-131 by the thyroid glands of patients and for monitoring the activities of I-131 in the thyroids of staff members who handle large activities of I-131. A thyroid probe, as shown in Figs. 20-23 and 20-24, usually consists of a 5.1-cm (2-inch) diameter and 5.1-cm-thick cylindrical NaI(Tl) crystal coupled to a PMT which in turn is connected to a preamplifier. The probe is shielded on the sides and back with lead and is equipped with a collimator so that it detects photons only from a limited portion of the patient. The thyroid probe is connected to a high-voltage power supply and either an SCA or an MCA system.

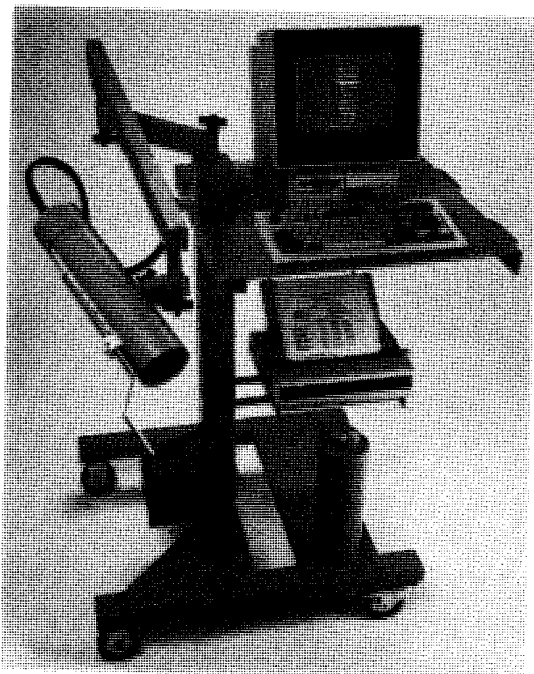


FIGURE 20-23. Thyroid probe system. (Photograph courtesy Capintec, Inc., Ramsey, New Jersey.) The personal computer has added circuitry and software so that it functions as a multichannel analyzer.

Thyroid Uptake Measurements

Thyroid uptake measurements may be performed using one or two capsules of I-123 or I-131 sodium iodide. A neck phantom, consisting of a Lucite cylinder of diameter similar to the neck and containing a hole parallel to its axis for a radioiodine capsule, is required. In the two-capsule method, the capsules should have almost identical activities. Each capsule is placed in the neck phantom and counted. Then, one capsule is swallowed by the patient. The other capsule is called the “standard.” Next, the emissions from the patient’s neck are counted, typically at 4 to 6 hours after administration, and again at 24 hours after administration. Each time that the patient’s thyroid is counted, the patient’s distal thigh is also counted for the same length of time, to approximate nonthyroidal activity in the neck, and a background count is obtained. All counts are performed with the NaI crystal at the same

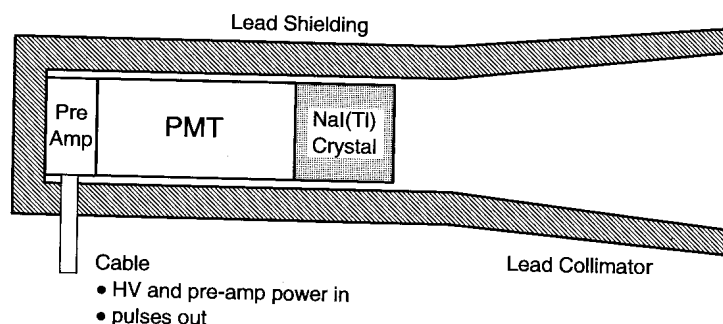


FIGURE 20-24. Diagram of a thyroid probe.

distance, typically 20 to 25 cm, from the thyroid phantom or the patient's neck or thigh. (This distance reduces the effects of small differences in distance between the detector and the objects being counted.) Furthermore, each time that the patient's thyroid is counted, the remaining capsule is placed in the neck phantom and counted. Finally, the uptake is calculated for each neck measurement:

$$\text{Uptake} = \frac{(\text{Thyroid count} - \text{Thigh count})}{(\text{Count of standard in phantom} - \text{Background count})} \times \frac{\text{Initial count of standard in phantom}}{\text{Initial count of patient capsule in phantom}}$$

Some nuclear medicine laboratories instead use a method that requires only one capsule. In this method, a single capsule is obtained, counted in the neck phantom, and swallowed by the patient. As in the previous method, the patient's neck and distal thigh are counted, typically at 4 to 6 hours and again at 24 hours after administration. The times of the capsule administration and the neck counts are recorded. Finally, the uptake is calculated for each neck measurement:

$$\frac{(\text{Thyroid count} - \text{Thigh count})}{(\text{Count of capsule in phantom} - \text{Background count})} \times e^{0.693t/T_{1/2}}$$

where $T_{1/2}$ is the physical half-life of the radionuclide and t is the time elapsed between the count of the capsule in the phantom and the thyroid count. The single-capsule method avoids the cost of the second capsule and requires fewer measurements, but it is more susceptible to instability of the equipment, technologist error, and dead-time effects.

Sodium Iodide Well Counter

Most nuclear medicine departments also have an NaI(Tl) well counter, shown in Fig. 20-25, for clinical tests such as Schilling tests (a test of vitamin B₁₂ absorption),

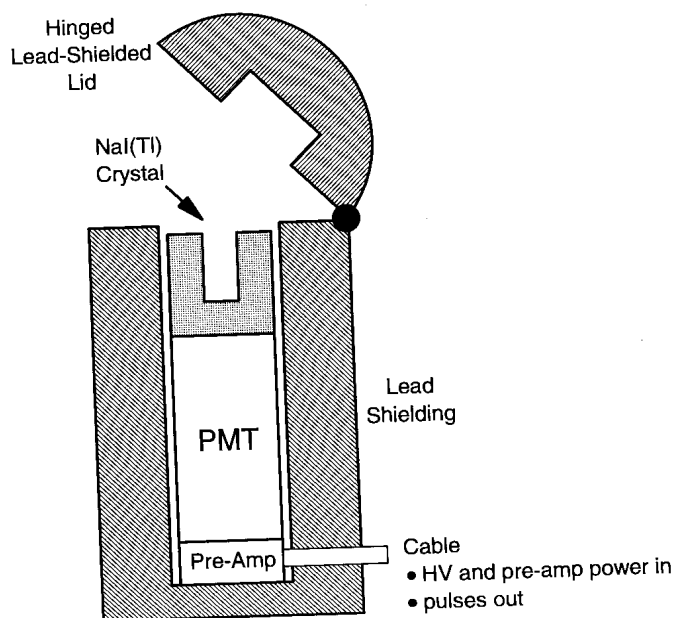


FIGURE 20-25. NaI(Tl) well counter.

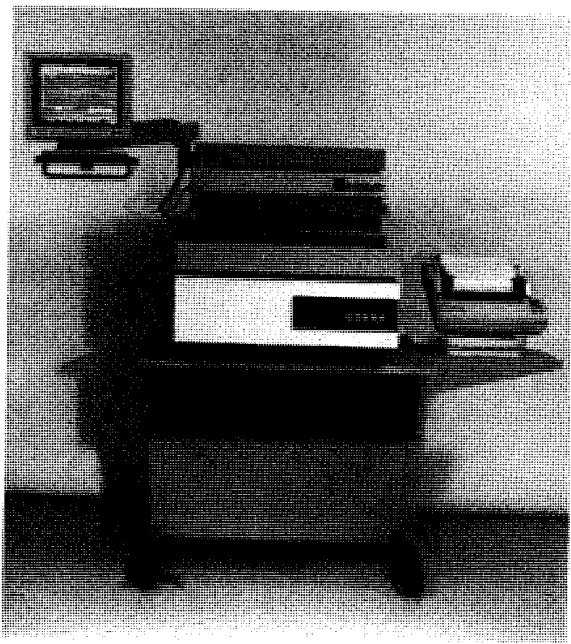


FIGURE 20-26. Automatic gamma well counter. (Photograph courtesy Packard Instrument Company.)

plasma or red blood cell volume determinations, and radioimmunoassays. The well counter usually consists of a cylindrical NaI(Tl) crystal, either 5.1 cm (2 inches) in diameter and 5.1 cm thick or 7.6 cm (3 inches) in diameter and 7.6 cm thick, with a hole in the crystal for the insertion of samples. This configuration gives the counter an extremely high efficiency, permitting it to assay samples containing activities of less than 1 nCi (10^{-3} μ Ci). The crystal is coupled to a PMT, which in turn is connected to a preamplifier. A well counter in a nuclear medicine department should have a thick lead shield, because it is used to count samples containing nanocurie activities in the vicinity of millicurie activities of high-energy gamma-ray emitters such as Ga-67, In-111, and I-131. The well counter is connected to a high-voltage power supply and either an SCA or an MCA system. Departments that perform large numbers of radioimmunoassays often use automatic well counters, such as the one shown in Fig. 20-26, to count large numbers of samples.

Sample Volume and Dead-Time Effects in Sodium Iodide Well Counters

The position of a sample in a sodium iodide well counter has a dramatic effect on the detection efficiency. When liquid samples in vials of a particular shape and size are counted, the detection efficiency falls as the volume increases. Most nuclear medicine *in vitro* tests require the comparison of a liquid sample from the patient with a reference sample. It is crucial that both samples be in identical containers and have identical volumes.

In addition, the high efficiency of the NaI well counter can cause unacceptable dead-time count losses, even with sample activities in the microcurie range. It is important to ensure that the activity placed in the well counter is sufficiently small that dead-time effects do not cause a falsely low count. In general, well counters should not be used at apparent count rates exceeding about 5,000 cps, which lim-

its samples of I-125 and Co-57 to activities less than about 0.2 μCi . However, larger activities of some radionuclides may be counted without significant losses; for example, activities of Cr-51 as large as 5 μCi may be counted, because only about one out of every ten decays yields a gamma ray.

Quality Assurance for the Sodium Iodide Thyroid Probe and Well Counter

Both of these instruments should have energy calibrations (as discussed earlier for an SCA system) performed daily, with the results recorded. A background count and a constancy test, using a source with a long half-life such as Cs-137, also should be performed daily for both the well counter and the thyroid probe to test for radioactive contamination or instrument malfunction. On the day the constancy test is begun, a counting window is set to tightly encompass the photopeak and a count is taken and corrected for background. Limits called action levels are established which, if exceeded, cause the individual performing the test to notify the chief technologist, physicist, or physician. On each subsequent day, a count is taken using the same source, window setting, and counting time; corrected for background; recorded; and compared with the action levels. (If each day's count were mistakenly compared with the previous day's count instead of the first day's count, slow changes in the instrument would not be discovered.) Periodically, the constancy test count will exceed the action levels because of source decay. When this happens, the first-day count and action levels should be corrected for decay.

Also, spectra should be plotted annually for commonly measured radionuclides, usually I-123 and I-131 for the thyroid probe and perhaps Co-57 (Schilling tests), I-125 (radioimmunoassays), and Cr-51 (red cell volumes and survival studies) for the well counter, to verify that the SCA windows fit the photopeaks. This testing is greatly simplified if the department has an MCA.

Dose Calibrator

A dose calibrator, shown in Fig. 20-27, is used to measure the activities of doses of radiopharmaceuticals to be administered to patients. The U.S. Nuclear Regulatory Commission (NRC) and state regulatory agencies require that doses of x-ray- and gamma ray-emitting radiopharmaceuticals be measured with a dose calibrator. Most dose calibrators are well-type ionization chambers that are filled with argon ($Z = 18$) and pressurized to maximize sensitivity. Some less expensive dose calibrators instead use GM tubes near a chamber for the insertion of the dose. Most dose calibrators have shielding around their chambers to protect users from the radioactive material being assayed and to prevent nearby sources of radiation from affecting the measurements.

A dose calibrator cannot directly measure activity. Instead, it measures the intensity of the radiation emitted by a dose of a radiopharmaceutical. The manufacturer of a dose calibrator determines calibration factors relating the intensity of the signal from the detector to activity for specific radionuclides commonly used in nuclear medicine. The user pushes a button or turns a dial on the dose calibrator to designate the radionuclide being measured, thereby specifying a calibration factor, and the dose calibrator displays the measured activity.

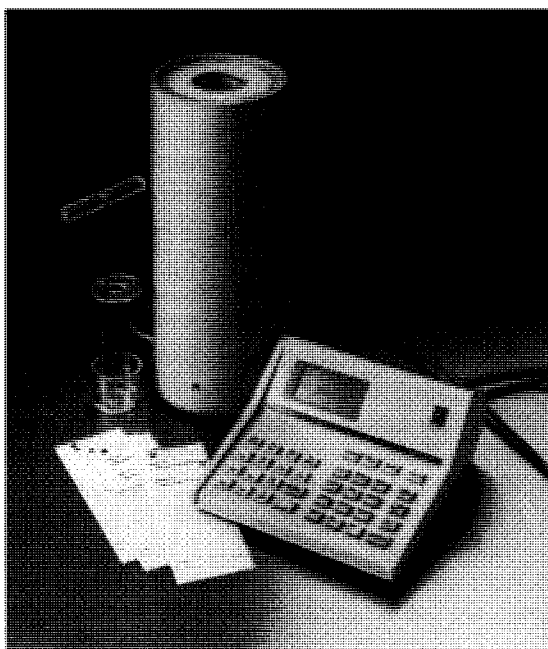


FIGURE 20-27. Dose calibrator. The detector is a well-type ion chamber filled with pressurized argon. The plastic vial and syringe holder (shown to the left of the ionization chamber) is used to place radioactive material into the detector. This reduces exposure to the hands and permits the activity to be measured in a reproducible geometry. (Photograph courtesy Capintec, Inc., Ramsey, New Jersey.)

Operating Characteristics

Dose calibrators using ionization chambers are operated in current mode, thereby avoiding dead-time effects. They can accurately assay activities as large as 2 Ci. For the same reasons, they are relatively insensitive and cannot accurately assay activities less than about 1 μ Ci. In general, the identification and measurement of activities of radionuclides in samples containing multiple radionuclides is not possible. The measurement accuracy is affected by the position in the well of the doses being measured, so it is important that all measurements be made with the doses at the same position. Most dose calibrators have large wells, which reduces the effect of position on the measurement.

Dose calibrators using large well-type ionization chambers are in general not significantly affected by changes in the sample volume or container for most radionuclides. However, the measured activities of certain radionuclides, especially those such as In-111, I-123, I-125, and Xe-133 that emit weak x-rays or gamma rays, are highly dependent on factors such as whether the container (i.e., syringe or vial) is glass or plastic and the thickness of the container's wall. There is currently no generally accepted solution to this problem. Some radiopharmaceutical manufacturers provide correction factors for these radionuclides; these correction factors are specific to the radionuclide, the container, and the model of dose calibrator.

There are even greater problems with attenuation effects when assaying the doses of pure beta-emitting radionuclides, such as P-32 and Sr-89. For this reason, the NRC does not require unit doses of these radionuclides to be assayed in a dose calibrator before administration. However, most nuclear medicine departments assay these as well, but only to verify that the activities as assayed by the vendor are not in error by a large amount.

Dose Calibrator Quality Assurance

Because the assay of activity using the dose calibrator is often the only assurance that the patient is receiving the prescribed activity, elaborate quality assurance testing of the dose calibrator is required by the NRC and state regulatory agencies. The device must be tested for *accuracy* on installation and annually thereafter. Two or more sealed radioactive sources (often Co-57 and Cs-137) whose activities are known within 5% are assayed. The measured activities must be within 10% of the actual activities.

The device also must be tested for *linearity* (a measure of the effect that the amount of activity has on the accuracy) on installation and quarterly thereafter. The most common method requires a vial of ^{99m}Tc containing the maximal activity that would be administered to a patient. The vial is measured two or three times daily until it contains less than 30 μCi ; the measured activities and times of the measurements are recorded. One measurement is assumed to be correct and, from this measurement, activities are calculated for the times of the other measurements. No measurement may differ from the corresponding calculated activity by more than 10%.

The device must be tested for *constancy* before its first use each day. At least one sealed source (usually Co-57) is assayed, and its measured activity, corrected for decay, must not differ from its measured activity on the date of the last accuracy test by more than 10%. (Most laboratories perform a daily accuracy test, which is more rigorous, in lieu of a daily constancy test.)

Finally, the dose calibrator is required to be tested for geometry dependence on installation. This is usually done by placing a small volume of a radiochemical, often ^{99m}Tc pertechnetate, in a vial or syringe and assaying its activity after each of several dilutions. If volume effects are found to affect measurements by more than 10%, correction factors must be determined. The geometry test must be performed for syringe and vial sizes commonly used. Dose calibrators must also be appropriately tested after repair or adjustment.

Although not required by regulatory agencies, the calibration factors of individual radionuclide settings should be verified for radionuclides assayed for clinical purposes. This can be performed by placing a source of any radionuclide in the dose calibrator, recording the indicated activity using a clinical radionuclide setting (e.g., I-131), recording the indicated activity using the setting of a radionuclide used for accuracy determination (e.g., Co-57 or Cs-137), and verifying that the ratio of the two indicated activities is that specified by the manufacturer of the dose calibrator.

Molybdenum Concentration Testing

When a molybdenum 99/technetium 99m (Mo-99/Tc-99m) generator is eluted, it is possible to obtain an abnormally large amount of Mo-99 in the eluate. (These generators are discussed in Chapter 19.) If a radiopharmaceutical contaminated with Mo-99 is administered to a patient, the patient will receive an increased radiation dose (Mo-99 emits high-energy beta particles and has a 66-hour half-life) and the quality of the resultant images will be degraded by the high-energy Mo-99 gamma rays. The NRC requires that any Tc-99m to be administered to a human must not contain more than 0.15 μCi of Mo-99 per mCi of Tc-99m at the time of administration and that each elution of a generator be assayed for Mo-99 concentration.

The concentration of Mo-99 is most commonly measured with a dose calibrator and a special lead container that is supplied by the manufacturer. The walls of the lead container are sufficiently thick to stop essentially all of the gamma rays from Tc-99m (140 keV) but thin enough to be penetrated by many of the higher-energy gamma rays from Mo-99 (740 and 778 keV). To perform the measurement, the empty lead container is first assayed in the dose calibrator. Next, the vial of Tc-99m is placed in the lead container and assayed. Finally, the vial of Tc-99m alone is assayed. The Mo-99 concentration is then obtained using the following equation:

$$\text{Concentration} = K \cdot \frac{A_{\text{vial-in-container}} - A_{\text{empty-container}}}{A_{\text{vial}}} \quad [20-5]$$

where K is a correction factor supplied by the manufacturer of the dose calibrator that accounts for the attenuation of the Mo-99 gamma rays by the lead container.

20.7 COUNTING STATISTICS

Introduction

Sources of Error

There are three types of errors in measurements. The first is *systematic error*. Systematic error occurs when measurements differ from the correct values in a systematic fashion. For example, systematic error occurs in radiation measurements if a detector is used in pulse mode at too high an interaction rate; dead-time count losses cause the measured count rate to be lower than the actual interaction rate. The second type of error is *random error*. Random error is caused by random fluctuations in whatever is being measured or in the measurement process itself. The third type of error is the *blunder* (e.g., seating the single channel analyzer window incorrectly for a single measurement).

Random Error in Radiation Detection

The processes by which radiation is emitted and interacts with matter are random in nature. Whether a particular radioactive nucleus decays within a specified time interval, the direction of an x-ray emitted by an electron striking the target of an x-ray tube, whether a particular x-ray passes through a patient to reach the film cassette of an x-ray machine, and whether a gamma ray incident upon a scintillation camera crystal is detected are all random phenomena. Therefore, all radiation measurements, including medical imaging, are subject to random error. Counting statistics enable judgments of the validity of measurements that are subject to random error.

Characterization of Data

Accuracy and Precision

If a measurement is close to the correct value, it is said to be *accurate*. If measurements are reproducible, they are said to be *precise*. Precision does not imply accuracy; a set of measurements may be very close together (precise) but not close to the correct value (i.e., inaccurate). If a set of measurements differs from the correct value in a systematic fashion (systematic error), the data are said to be *biased*.

Measures of Central Tendency—Mean and Median

Two measures of the central tendency of a set of measurements are the *mean* (average) and the *median*. The mean ($\bar{x}$) of a set of measurements is defined as follows:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad [20-6]$$

where N is the number of measurements.

To obtain the *median* of a set of measurements, they must first be sorted by size. The median is the middlemost measurement if the number of measurements is odd, and it is the average of the two middlemost measurements if the number of measurements is even. For example, to obtain the median of the five measurements 8, 14, 5, 9, and 12, they are first sorted by size: 5, 8, 9, 12, and 14. The median is 9. The advantage of the median over the mean is that the median is less affected by outliers. An outlier is a measurement that is much greater or much less than the others.

Measures of Variability—Variance and Standard Deviation

The variance and standard deviation are measures of the variability (spread) of a set of measurements. The variance (σ^2) is determined from a set of measurements by subtracting the mean from each measurement, squaring the difference, summing the squares, and dividing by one less than the number of measurements:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N - 1} \quad [20-7]$$

where N is the total number of measurements and $\bar{x}$ is the sample mean.

The *standard deviation* (σ) is the square-root of the variance:

$$\sigma = \sqrt{\sigma^2} \quad [20-8]$$

The *fractional standard deviation* (also referred to as the *fractional error* or *coefficient of variation*) is the standard deviation divided by the mean:

$$\text{Fractional standard deviation} = \sigma/\bar{x} \quad [20-9]$$

Probability Distribution Functions for Binary Processes

Binary Processes

A *trial* is an event that may have more than one outcome. A *binary process* is a process in which a trial can only have two outcomes, one of which is arbitrarily called a *success*. A toss of a coin is a binary process. The toss of a die can be considered a binary process if, for example, a “two” is selected as a success and all other outcomes are considered to be failures. Whether a particular radioactive nucleus decays during a specified time interval is a binary process. Whether a particular x-ray or gamma ray is detected by a radiation detector is a binary process. Table 20-2 lists examples of binary processes.

A measurement consists of counting the number of successes from a specified number of trials. Tossing ten coins and counting the number of “heads” is a measurement. Placing a radioactive sample on a detector and recording the number of events detected is a measurement.

TABLE 20-2. BINARY PROCESSES

Trial	Definition of a success	Probability of a success
Toss of a coin	"Heads"	1/2
Toss of a die	"A four"	1/6
Observation of a radioactive nucleus for a time " t "	It decays	$1 - e^{-\lambda t}$
Observation of a detector of efficiency E placed near a radioactive nucleus for a time " t "	A count	$E(1 - e^{-\lambda t})$

Source: Adapted from Knoll, GF. *Radiation detection and measurement*, 3rd ed. New York: John Wiley, 2000.

Probability Distribution Functions—Binomial, Poisson, and Gaussian

A probability distribution function (pdf) (also known as the probability density function) describes the probability of obtaining each outcome from a measurement—for example, the probability of obtaining six "heads" in a throw of ten coins. There are three pdfs relevant to binary processes—the binomial, the Poisson, and the Gaussian (normal). The binomial pdf exactly describes the probability of each outcome from a measurement of a binary process:

$$P(x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \quad [20-10]$$

where N is the total number of trials in a measurement, p is the probability of success in a single trial, and x is the number of successes. The mathematical notation $N!$, called *factorial notation*, is simply shorthand for the product

$$N! = N \cdot (N-1) \cdot (N-2) \dots 3 \cdot 2 \cdot 1 \quad [20-11]$$

For example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. If we wish to know the probability of obtaining two heads in a toss of four coins, $x = 2$, $N = 4$, and $p = 0.5$. The probability of obtaining two heads is

$$P(\text{two-heads}) = \frac{4!}{2!(4-2)!} (0.5)^2 (1-0.5)^{4-2} = 0.375$$

Figure 20-28 is a graph of the binomial pdf. It can be shown that the sum of the probabilities of all outcomes for the binomial pdf is 1.0 and that the mean ($\bar{x}$) and standard deviation (σ) of the binomial pdf are as follows:

$$\bar{x} = pN \quad \text{and} \quad \sigma = \sqrt{pN(1-p)} \quad [20-12]$$

If the probability of a success in a trial is much less than 1 (not true for a toss of a coin, but true for most radiation measurements), the standard deviation is approximated by the following:

$$\sigma = \sqrt{pN(1-p)} \approx \sqrt{pN} = \sqrt{\bar{x}} \quad [20-13]$$

Because of the factorials in Equation 20-10, it is difficult to use if either x or N is large. The Poisson and Gaussian pdfs are approximations to the binomial pdf that are often used when x or N is large. Figure 20-29 shows a Gaussian pdf.

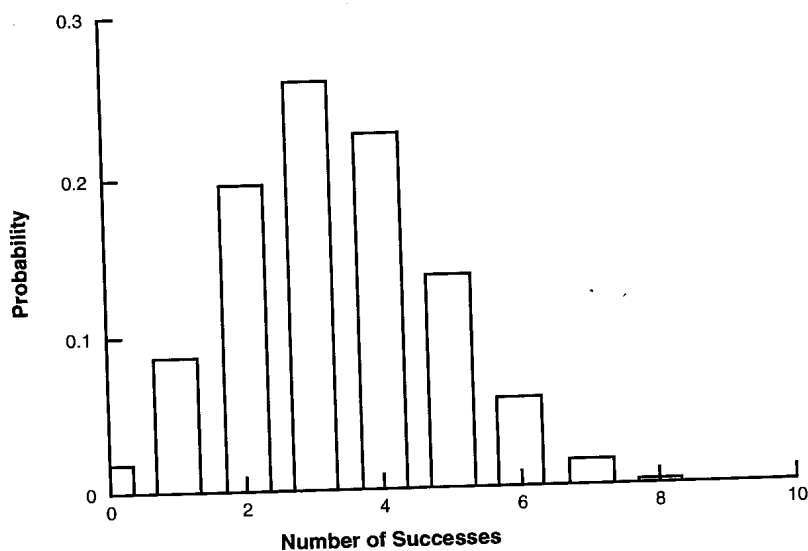


FIGURE 20-28. Binomial probability distribution function when the probability of a success in a single trial (p) is $1/3$ and the number of trials (N) is 10.

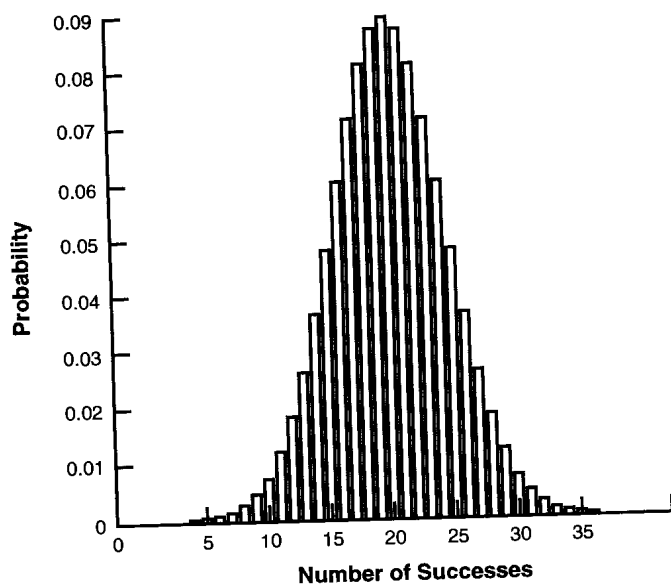


FIGURE 20-29. Gaussian probability distribution function for $pN = 20$.

Estimating the Uncertainty of a Single Measurement

Estimated Standard Deviation

The standard deviation can be estimated, as previously described, by making several measurements and applying Equations 20-7 and 20-8. Nevertheless, if the process being measured is a binary process, the standard deviation can be estimated from a single measurement. The single measurement is probably close to the mean.

**TABLE 20-3. FRACTIONAL ERRORS
(PERCENT STANDARD DEVIATIONS)
FOR SEVERAL NUMBERS OF COUNTS**

Count	Fractional error (%)
100	10.0
1,000	3.2
10,000	1.0
100,000	0.32

Because the standard deviation is approximately the square-root of the mean, it is also approximately the square-root of the single measurement:

$$\sigma \approx \sqrt{x} \quad [20-14]$$

where x is the single measurement. The fractional standard deviation of a single measurement may also be estimated:

$$\text{Fractional error} = \frac{\sigma}{x} \approx \frac{\sqrt{x}}{x} = \frac{1}{\sqrt{x}} \quad [20-15]$$

For example, a single measurement of a radioactive source yields 1,256 counts. The estimated standard deviation is:

$$\sigma \approx \sqrt{1,256 \text{ cts}} = 35.4 \text{ cts}$$

The fractional standard deviation is estimated as:

$$\text{Fractional error} = 35.4 \text{ cts} / 1,256 \text{ cts} = 0.028 = 2.8 \%$$

Table 20-3 lists the estimated fractional errors for various numbers of counts. The fractional error decreases with the number of counts.

Confidence Intervals

Table 20-4 lists intervals about a measurement, called *confidence intervals*, for which the probability of containing the true mean is specified. There is a 68.3% probability that the true mean is within one standard deviation (1σ) of a measurement, a 95% probability that it is within 2σ of a measurement, and a 99.7% probability that it is within 3σ of a measurement, assuming that the measurements follow a Gaussian distribution.

TABLE 20-4. CONFIDENCE INTERVALS

Interval about measurement	Probability that mean is within interval (%)
$\pm 0.674\sigma$	50.0
$\pm 1\sigma$	68.3
$\pm 1.64\sigma$	90.0
$\pm 1.96\sigma$	95.0
$\pm 2.58\sigma$	99.0
$\pm 3\sigma$	99.7

Source: Adapted from Knoll, GF. *Radiation detection and measurement*, 2nd ed. New York: John Wiley, 1989.

For example, a count of 853 is obtained. Determine the interval about this count in which there is a 95% probability of finding the true mean.

First, the standard deviation is estimated:

$$\sigma \approx \sqrt{853 \text{ cts}} = 29.2 \text{ cts}$$

From Table 20-4, the 95% confidence interval is determined as follows:

$$853 \text{ cts} \pm 1.96 \sigma = 853 \text{ cts} \pm 1.96 (29.2 \text{ cts}) = 853 \text{ cts} \pm 57.2 \text{ cts}$$

So, the 95% confidence interval ranges from 796 to 910 counts.

Propagation of Error

In nuclear medicine, calculations are frequently performed using numbers that incorporate random error. It is often necessary to estimate the uncertainty in the results of these calculations. Although the standard deviation of a count may be estimated by simply taking its square root, it is incorrect to calculate the standard deviation of the result of a calculation by taking its square root. Instead, the standard deviations of the actual counts must first be calculated and entered into propagation of error equations to obtain the standard deviation of the result.

Multiplication or Division of a Number with Error by a Number without Error

It is often necessary to multiply or divide a number containing random error by a number that does not contain random error. For example, to calculate a count rate, a count (which incorporates random error) is divided by a counting time (which does involve significant random error). If a number x has a standard deviation σ and is multiplied by a number c without random error, the standard deviation of the product cx is $c\sigma$. If a number x has a standard deviation σ and is divided by a number c without random error, the standard deviation of the quotient x/c is σ/c .

For example, a 5-minute count of a radioactive sample yields 952 counts. The count rate is $952 \text{ cts}/5 \text{ min} = 190 \text{ cts/min}$. The standard deviation and percent standard deviation of the count are:

$$\sigma \approx \sqrt{952 \text{ cts}} = 30.8 \text{ cts}$$

$$\text{Fractional error} = 30.8 \text{ cts}/952 \text{ cts} = 0.032 = 3.2\%$$

The standard deviation of the count rate is:

$$\sigma = 30.8 \text{ cts} / 5 \text{ min} = 6.16 \text{ cts/min}$$

$$\text{Fractional error} = 6.16 \text{ cts}/190 \text{ cts} = 3.2\%$$

Notice that the percent standard deviation is not affected when a number is multiplied or divided by a number without random error.

Addition or Subtraction of Numbers with Error

It is often necessary to add or subtract numbers with random error. For example, a background count may be subtracted from a count of a radioactive sample.

TABLE 20-5. PROPAGATION OF ERROR EQUATIONS^a

Description	Operation	Standard deviation
Multiplication of a number with random error by a number without random error	cx	$c\sigma$
Division of a number with random error by a number without random error	x/c	σ/c
Addition of two numbers containing random errors	$x_1 + x_2$	$\sqrt{\sigma_1^2 + \sigma_2^2}$
Subtraction of two numbers containing random errors	$x_1 - x_2$	$\sqrt{\sigma_1^2 + \sigma_2^2}$

^a c is a number without random error, σ is the standard deviation of x , σ_1 is the standard deviation of x_1 , and σ_2 is the standard deviation of x_2 .

Whether two numbers are added or subtracted, the same equation is used to calculate the standard deviation of the result, as shown in Table 20-5.

For example, a count of a radioactive sample yields 1,952 counts, and a background count with the sample removed from the detector yields 1,451 counts.

The count of the sample, corrected for background, is:

$$1,952 \text{ cts} - 1,451 \text{ cts} = 501 \text{ cts}$$

The standard deviation and percent standard deviation of the original sample count are:

$$\sigma_{s+b} \approx \sqrt{1,952 \text{ cts}} = 44.2 \text{ cts}$$

$$\text{Fractional error} = 44.2 \text{ cts} / 1,952 \text{ cts} = 2.3\%$$

The standard deviation and percent standard deviation of the background count are:

$$\sigma_b \approx \sqrt{1,451 \text{ cts}} = 38.1 \text{ cts}$$

$$\text{Fractional error} = 38.1 \text{ cts} / 1,451 \text{ cts} = 2.6\%$$

The standard deviation and percent standard deviation of the sample count corrected for background are:

$$\sigma_s \approx \sqrt{(44.2 \text{ cts})^2 + (38.3 \text{ cts})^2} = 58.3 \text{ cts}$$

$$\text{Fractional error} = 58.1 \text{ cts} / 501 \text{ cts} = 11.6\%$$

Note that the fractional error of the difference is much larger than the fractional error of either count. The fractional error of the difference of two numbers of similar magnitude can be much larger than the fractional errors of the two numbers.

Combination Problems

It is sometimes necessary to perform mathematical operations in series, for example, to subtract two numbers and then to divide the difference by another number. The standard deviation is calculated for each intermediate result in the series by

entering the standard deviations from the previous step into the appropriate propagation of error equation in Table 20-5.

For example, a 5-minute count of a radioactive sample yields 1,290 counts and a 5-minute background count yields 451 counts. What is the count-rate due to the sample alone and its standard deviation?

First, the count is corrected for background:

$$\text{Count} = 1,290 \text{ cts} - 451 \text{ cts} = 839 \text{ cts}$$

The estimated standard deviation of each count and the difference are calculated:

$$\sigma_{s+b} \approx \sqrt{1,290 \text{ cts}} = 35.9 \text{ cts} \quad \text{and} \quad \sigma_b \approx \sqrt{451 \text{ cts}} = 21.2 \text{ cts}$$

$$\sigma_s \approx \sqrt{(35.9 \text{ cts})^2 + (21.2 \text{ cts})^2} = 41.7 \text{ cts}$$

Finally, the count rate due to the source alone and its standard deviation are calculated:

$$\text{Count rate} = 839 \text{ cts}/5 \text{ min} = 168 \text{ cts/min}$$

$$\sigma_{s+b}/c = 41.7 \text{ cts}/5 \text{ min} = 8.3 \text{ cts/min}$$

SUGGESTED READING

- Knoll GF. *Radiation detection and measurement*, 3rd ed. New York: John Wiley, 2000.
- Patton JA, Harris CC. Gamma well counter. In: Sandler MP, et al., eds. *Diagnostic nuclear medicine*, 3rd ed. Baltimore: Williams & Wilkins, 1996:59–65.
- Ranger NT. The AAPM/RSNA physics tutorial for residents: radiation detectors in nuclear medicine. *Radiographics* 1999;19:481–502.
- Ranger NT. Radiation detectors in nuclear medicine. *Radiographics* 1999;19:481–502.
- Rzeszutowski MS. The AAPM/RSNA physics tutorial for residents: counting statistics. *Radiographics* 1999;19:765–782.
- Sorenson JA, Phelps ME. *Physics in nuclear medicine*, 2nd ed. New York: Grune and Stratton, 1987.

NUCLEAR IMAGING—THE SCINTILLATION CAMERA

Nuclear imaging produces images of the distributions of radionuclides in patients. Because charged particles from radioactivity in a patient are almost entirely absorbed within the patient, nuclear imaging uses gamma rays, characteristic x-rays (usually from radionuclides that decay by electron capture), or annihilation photons (from positron-emitting radionuclides) to form images.

To form a projection image, an imaging system must determine not only the photon flux density (number of x- or gamma rays per unit area) at each point in the image plane but also the directions of the detected photons. In x-ray transmission imaging, the primary photons travel known paths diverging radially from a point (the focal spot of the x-ray tube). In contrast, the x- or gamma rays from the radionuclide at each portion of a patient are emitted isotropically (equally in all directions). Nuclear medicine instruments designed to image gamma ray and x-ray emitting radionuclides use *collimators* that permit photons following certain trajectories to reach the detector but absorb most of the rest. A heavy price is paid for using collimation—the vast majority (typically well over 99.95%) of emitted photons is wasted. Thus collimation, although essential to image formation, severely limits the performance of these devices. Instruments for imaging positron (β^+) emitting radionuclides can avoid collimation by exploiting the unique properties of annihilation radiation to determine the directions of the photons.

The earliest successful nuclear medicine imaging device, the rectilinear scanner, which dominated nuclear imaging from the early 1950s through the late 1960s, used a single moving radiation detector to sample the photon fluence at a small region of the image plane at a time. This was improved upon by the use of a large-area position-sensitive detector (a detector indicating the location of each interaction) to sample simultaneously the photon fluence over the entire image plane. The Anger scintillation camera, which currently dominates nuclear imaging, is an example of the latter method. The scanning detector system is less expensive, but the position-sensitive detector system permits more rapid image acquisition and has replaced single scanning detector systems.

Nuclear imaging devices using gas-filled detectors (such as multiwire proportional counters) have been developed. Unfortunately, the low densities of gases, even when pressurized, yield low detection efficiencies for the x- and gamma ray energies commonly used in nuclear imaging. To obtain a sufficient number of interactions to form statistically valid images without imparting an excessive radiation

dose to the patient, nearly all nuclear imaging devices in routine clinical use utilize solid inorganic scintillators as detectors because of their superior detection efficiency. Efforts are being made to develop nuclear imaging devices using semiconductor detectors. This will require the development of a high atomic number, high-density semiconductor detector of sufficient thickness for absorbing the x- and gamma rays commonly used in nuclear imaging and that can be operated at room temperature. The leading detector material to date is cadmium zinc telluride (CZT). A small field-of-view camera using CZT semiconductor detectors has been developed, and scintillation camera manufacturers are investigating this material.

The attenuation of x-rays in the patient is useful in radiography and fluoroscopy, and, in fact, is necessary for image formation. However, in nuclear imaging, attenuation is usually a hindrance; it causes a loss of information and, especially when it is very nonuniform, it is a source of artifacts.

The quality of a nuclear medicine image is determined not only by the performance of the nuclear imaging device but also by the properties of the radiopharmaceutical used. For example, the ability of a radiopharmaceutical to preferentially accumulate in a lesion of interest largely determines the smallest such lesion that can be detected. Furthermore, the dosimetric properties of a radiopharmaceutical determine the maximal activity that can be administered to a patient, which in turn affects the amount of statistical noise (quantum mottle) in the image and the spatial resolution. For example, in the 1950s and 1960s, radiopharmaceuticals labeled with I-131 and Hg-203 were commonly used. The long half-lives of these radionuclides and their beta particle emissions limited administered activities to approximately 150 microcuries (μCi). These low administered activities and the high-energy gamma rays of these radionuclides required the use of collimators providing poor spatial resolution. Many modern radiopharmaceuticals are labeled with technetium-99m (Tc-99m), whose short half-life (6.01 hours) and isomeric transition decay (~88% of the energy is emitted as 140 keV gamma rays; only ~12% is given to conversion electrons and other emissions unlikely to escape the patient) permit activities of up to 35 millicuries (mCi) to be administered. The high gamma-ray flux from such an activity permits the use of high-resolution (low-efficiency) collimators. Radiopharmaceuticals are discussed in Chapter 19.

21.1 PLANAR NUCLEAR IMAGING: THE ANGER SCINTILLATION CAMERA

The Anger gamma scintillation camera, developed by Hal O. Anger at the Donner Laboratory in Berkeley, California, in the 1950s, is by far the most common nuclear medicine imaging device. However, it did not begin to replace the rectilinear scanner significantly until the late 1960s, when its spatial resolution became comparable to that of the rectilinear scanner and Tc-99m-labeled radiopharmaceuticals, for which it is ideally suited, became commonly used in nuclear medicine. Most of the advantages of the scintillation camera over the rectilinear scanner stem from its ability simultaneously to collect data over a large area of the patient, rather than one small area at a time. This permits the more rapid acquisition of images and enables dynamic studies that depict the redistribution of radionuclides to be performed. Because the scintillation camera wastes fewer x- or gamma rays than earlier imaging devices, its images have less quantum mottle (statistical noise) and it can be used

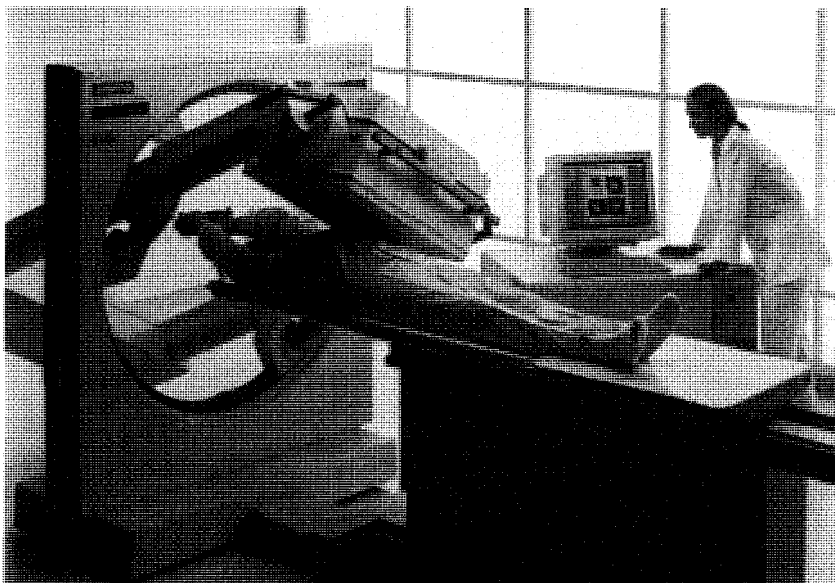


FIGURE 21-1. Modern rectangular head, large field-of-view scintillation camera. (Courtesy of Siemens Medical Systems, Nuclear Medicine Group.)

with higher resolution collimators, thus producing images of better spatial resolution. The scintillation camera is also more flexible in its positioning, permitting images to be obtained from almost any angle. Although it can produce satisfactory images using x- or gamma rays ranging in energy from about 70 keV (Tl-201) to 364 keV (I-131) or even 511 keV (F-18), the scintillation camera is best suited for imaging photons with energies in the range of 100 to 200 keV. Figure 21-1 shows a modern scintillation camera.

Scintillation cameras of other designs have been devised, and one, the multicrystal scintillation camera, achieved limited commercial success. However, the superior performance of the Anger camera for most applications has caused it to dominate nuclear imaging. The term *scintillation camera* will refer exclusively to the Anger scintillation camera throughout this chapter.

Design and Principles of Operation

Detector and Electronic Circuits

A scintillation camera (Fig. 21-2), contains a disk-shaped or rectangular thallium-activated sodium iodide NaI(Tl) crystal, typically 0.95 cm ($\frac{3}{8}$ inch) thick, optically coupled to a large number (typically 37 to 91) of 5.1- to 7.6-cm (2- to 3-inch) diameter photomultiplier tubes (PMTs). PMTs were described in Chapter 20. Some camera designs incorporate a Lucite light-pipe between the glass cover of the crystal and PMTs; in others, the PMTs are directly coupled to the glass cover. In most cameras, a preamplifier is connected to the output of each PMT. Between the patient and the crystal is a collimator, usually made of lead, that only allows x- or gamma rays approaching from certain directions to reach the crystal. The collima-

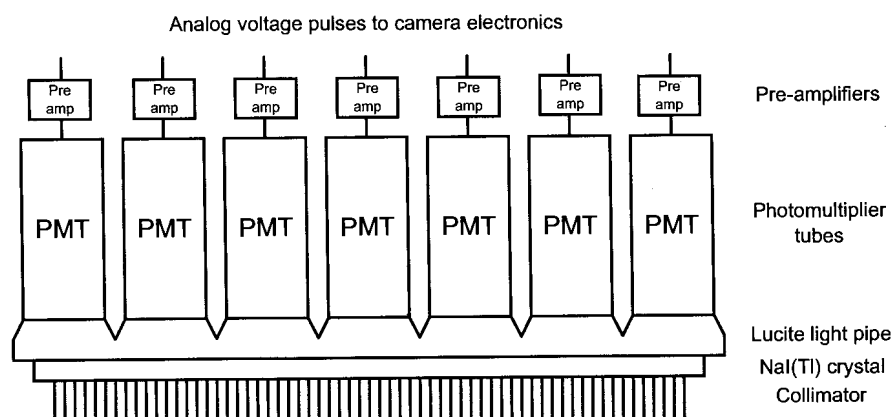


FIGURE 21-2. Scintillation camera detector.

tor is essential; a scintillation camera without a collimator does not generate meaningful images. Figure 21-2 shows a parallel-hole collimator.

The lead walls, called *septa*, between the holes in the collimator absorb most photons approaching the collimator from directions that are not aligned with the holes. Most photons approaching the collimator from a nearly perpendicular direction pass through the holes; many of these are absorbed in the sodium iodide crystal, causing the emission of visible and ultraviolet light. The light photons are converted into electrical signals and amplified by the PMTs. These signals are further amplified by the preamplifiers (preamps). The amplitude of the pulse produced by each PMT is proportional to the amount of light it received from an x- or gamma-ray interaction in the crystal.

Because the collimator septa intercept most photons approaching the camera from nonperpendicular directions, the pattern of photon interactions in the crystal forms a two-dimensional projection of the three-dimensional activity distribution in the patient. The PMTs closest to each photon interaction in the crystal receive more light than those that are more distant, causing them to produce larger voltage pulses. The relative amplitudes of the pulses from the PMTs following each interaction contain sufficient information to determine the location of the interaction in the plane of the crystal.

Until the late 1970s, scintillation cameras formed images as shown in Fig. 21-3, using only analog circuitry. The pulses from the preamps were sent to two circuits. The position circuit received the pulses from the individual preamps after each x- or gamma-ray interaction in the crystal and, by determining the centroid of these pulses, produced an X-position pulse and a Y-position pulse that together specified the position of the interaction in the plane of the crystal. The summing circuit added the pulses from the individual preamps to produce an energy (Z) pulse proportional in amplitude to the total energy deposited in the crystal. The Z pulse from the summing circuit was sent to a single-channel analyzer (SCA). The SCA, described in Chapter 20, produced a logic pulse (a voltage pulse of fixed amplitude and duration) only if the Z pulse received was within a preset range of energies. An interaction in the camera's crystal was recorded in the image only if a logic pulse was

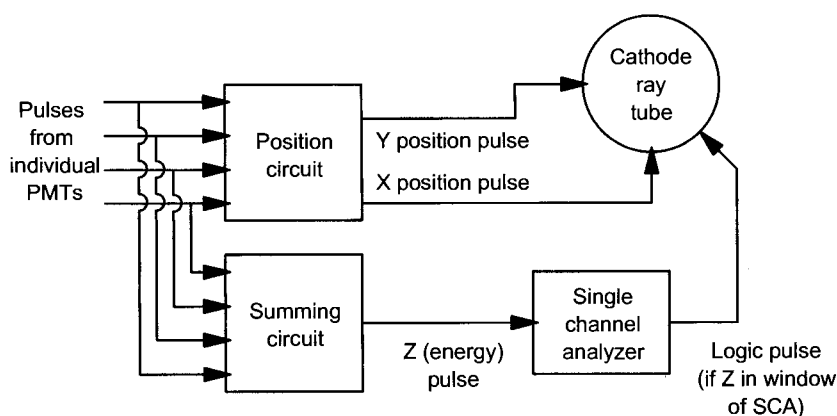


FIGURE 21-3. Electronic circuits of a fully analog scintillation camera.

produced by the SCA. These analog scintillation cameras had two to four SCAs for imaging radionuclides, such as Ga-67 and In-111, which emit useful photons of more than one energy.

In these scintillation cameras, the analog X- and Y-position pulses and the logic pulse from each interaction were then sent to a cathode ray tube (CRT). CRTs are described in Chapter 4. The CRT produced a momentary dot of light on its screen at a position determined by the X and Y pulses if a logic pulse from a SCA was received simultaneously. A photographic camera aimed at the CRT recorded the flashes of light, forming an image on film, dot by dot.

In the late 1970s, hybrid analog-digital scintillation cameras were introduced (Fig. 21-4). Hybrid cameras have analog position and summing circuits, as did the

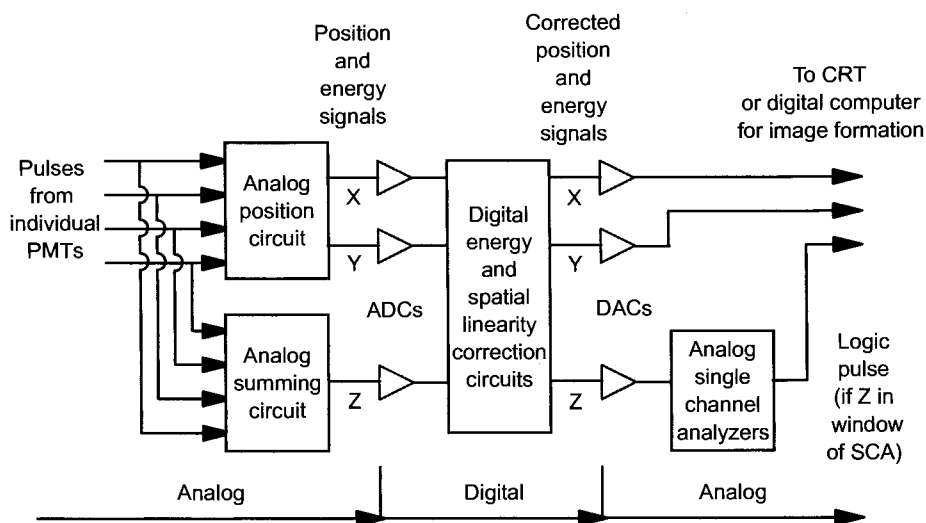


FIGURE 21-4. Electronic circuits of a hybrid analog-digital scintillation camera.

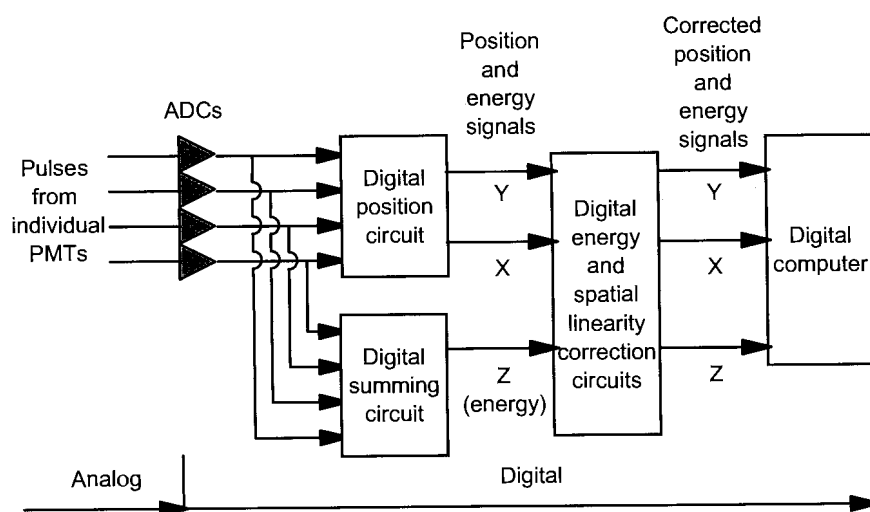


FIGURE 21-5. Electronic circuits of a modern digital scintillation camera.

earlier fully analog scintillation cameras. However, in hybrid cameras, the analog X, Y, and Z pulses from the position and summing circuits are converted to digital signals by analog-to-digital converters (ADCs). ADCs are described in Chapter 4. These digital signals are then sent to digital correction circuits, described later in this chapter (see Spatial Linearity and Uniformity, below). Following the digital correction circuits, the corrected digital X, Y, and Z signals are converted back to analog voltage pulses. Energy discrimination is performed in the analog domain by SCAs. Following energy discrimination, the X- and Y-position pulses can be sent to a CRT and recorded on photographic film, as described above. Alternatively, they can be sent to a digital computer, where they are digitized again and formed into a digital image, which can be displayed on a video monitor. Digital correction circuits greatly improve the spatial linearity and uniformity of the images. Many hybrid cameras are still in use today.

Today, the scintillation cameras made by several manufacturers provide an ADC for the signal from the preamplifier following each photomultiplier tube (Fig. 21-5). In these cameras, the position signals and usually the energy (Z) signals are formed using digital circuitry. These digital X, Y, and Z signals are corrected, using digital correction circuits, and energy discrimination is applied, also in the digital domain. These signals are then formed into digital images within a computer.

Collimators

The collimator of a scintillation camera forms the projection image by permitting x- or gamma-ray photons approaching the camera from certain directions to reach the crystal while absorbing most of the other photons. Collimators are made of high atomic number, high-density materials, usually lead. The most commonly used collimator is the *parallel-hole collimator*, which contains thousands of parallel holes (Fig. 21-6). The holes may be round, square, or triangular; however, most state-of-

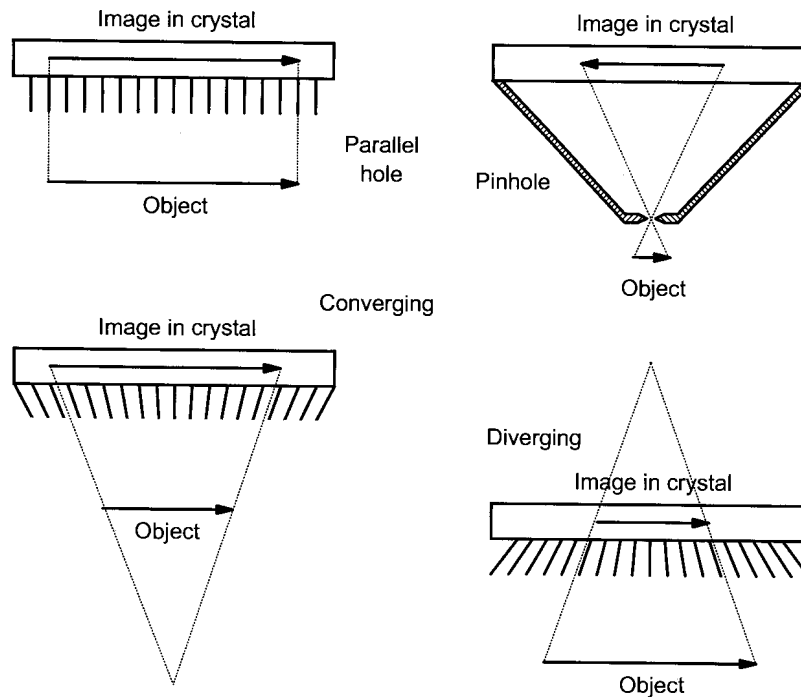


FIGURE 21-6. Collimators.

the-art collimators have hexagonal holes and are usually made from lead foil, although some are cast. The partitions between the holes are called septa. The septa must be thick enough to absorb most of the photons incident upon them. For this reason, collimators designed for use with radionuclides that emit higher-energy photons have thicker septa. There is an inherent compromise between the spatial resolution and efficiency (sensitivity) of collimators. Modifying a collimator to improve its spatial resolution (e.g., by reducing the size of the holes or lengthening the collimator) reduces its efficiency. Most scintillation cameras are provided with a selection of parallel-hole collimators. These may include "low-energy, high-sensitivity," "low-energy, all-purpose" (LEAP), "low-energy, high-resolution," "medium-energy" (suitable for Ga-67 and In-111), "high-energy" (for I-131), and "ultra-high-energy" (for F 18) collimators. The size of the image produced by a parallel-hole collimator is not affected by the distance of the object from the collimator. However, its spatial resolution degrades rapidly with increasing collimator-to-object distance.

A *pinhole collimator* (Fig. 21-6) is commonly used to produce magnified views of small objects, such as the thyroid or a hip joint. It consists of a small (typically 3- to 5-mm diameter) hole in a piece of lead or tungsten mounted at the apex of a leaded cone. Its function is identical to the pinhole in a pinhole photographic camera. As shown in the figure, the pinhole collimator produces a magnified image whose orientation is reversed. The magnification of the pinhole collimator decreases as an object is moved away from the pinhole. If an object is as far from the pinhole as the pinhole is from the crystal of the camera, the object is not magnified and, if

the object is moved yet farther from the pinhole, it is minified. (Clinical imaging is not performed at these distances.) There are pitfalls in the use of pinhole collimators due to the decreasing magnification with distance. For example, a thyroid nodule deep in the mediastinum can appear to be in the thyroid itself. Pinhole collimators are used extensively in pediatric nuclear medicine.

A *converging collimator* (Fig. 21-6) has many holes, all aimed at a focal point in front of the camera. As shown in Fig. 21-6, the converging collimator magnifies the image. The magnification increases as the object is moved away from the collimator. A *diverging collimator* (Fig. 21-6) has many holes aimed at a focal point behind the camera. It produces a minified image in which the amount of minification increases as the object is moved away from the camera. A diverging collimator may be used to image a large portion of a patient on a small (25-cm diameter) or standard (30-cm diameter) field-of-view (FOV) camera. For example, a diverging collimator would enable a mobile scintillation camera with a standard FOV to perform a lung study of a patient in the intensive care unit. If a diverging collimator is reversed on a camera, it becomes a converging collimator. The diverging collimator is seldom used today, as it has inferior imaging characteristics to the parallel-hole collimator, and the large crystal sizes of most modern cameras render it unnecessary. The converging collimator is also seldom used; its imaging characteristics are superior, in theory, to the parallel-hole collimator, but its decreasing FOV with distance and varying magnification with distance have discouraged its use. However, a hybrid of the parallel-hole and converging collimator, called a *fan-beam collimator*, is used in single photon emission computed tomography (SPECT) to take advantage of the favorable imaging properties of the converging collimator (see Chapter 22).

Many special-purpose collimators, such as the seven-pinhole collimator, have been developed. However, most of them have not enjoyed wide acceptance. The performance characteristics of parallel-hole, pinhole, and fan-beam collimators are discussed below (see Performance).

Principles of Image Formation

In nuclear imaging, which can also be called emission imaging, the photons from each point in the patient are emitted isotropically. Figure 21-7 shows the fates of the x- and gamma rays emitted in a patient being imaged. Some photons escape the patient without interaction, some scatter within the patient before escaping, and some are absorbed within the patient. Many of the photons escaping the patient are not detected because they are emitted in directions away from the detector. The collimator absorbs the vast majority of those photons that reach it. As a result, only a tiny fraction of the emitted photons (about one to two in 10,000 for typical low-energy parallel-hole collimators) has trajectories permitting passage through the collimator holes; thus, well over 99.9% of the photons emitted during imaging are wasted. Some photons penetrate the septa of the collimator without interaction. Of those reaching the crystal, some are absorbed in the crystal, some scatter from the crystal, and some pass through the crystal without interaction. The relative probabilities of these events depend on the energies of the photons and the thickness of the crystal. Of those photons absorbed in the crystal, some are absorbed by a single photoelectric absorption, whereas others undergo one or more Compton scatters before a photoelectric absorption. It is also possible

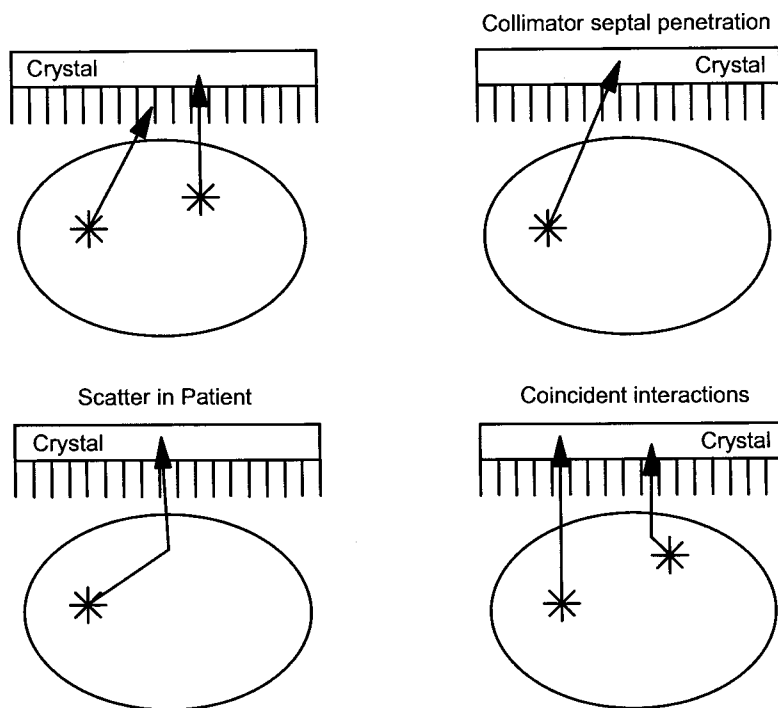


FIGURE 21-7. Ways that x- and gamma rays interact with a scintillation camera. All of these, other than the ones depicted in the upper left, cause a loss of contrast and spatial resolution. However, interactions by photons that have scattered through large angles and many coincident interactions are rejected by pulse height discrimination circuits.

for two photons simultaneously to interact with the crystal; if the energy (Z) signal from the coincident interactions is within the energy window of the energy discrimination circuit, the result will be a single count mispositioned in the image. The fraction of simultaneous interactions increases with the interaction rate of the camera.

Interactions in the crystal of photons that have been scattered in the patient, photons that have penetrated the collimator septa, photons that undergo one or more scatters in the crystal, and coincident interactions all reduce the spatial resolution and image contrast. The function of the camera's energy discrimination circuits (also known as pulse height analyzers) is to reduce this loss of resolution and contrast by rejecting photons that scatter in the patient or result in coincident interactions. Unfortunately, the limited energy resolution of the camera causes a wide photopeak, and low-energy photons can scatter through large angles with only a small energy loss. For example, a 140-keV photon scattering 45 degrees will only lose 7.4% of its energy. An energy window that encompasses most of the photopeak will unfortunately still accept a significant fraction of the scattered photons and coincident interactions.

It is instructive to compare single photon emission imaging with x-ray transmission imaging (Table 21-1). In x-ray transmission imaging, including radiog-

TABLE 21-1. COMPARISON OF SINGLE-PHOTON NUCLEAR IMAGING WITH X-RAY TRANSMISSION IMAGING

	Principle of image formation	Scatter rejection	Pulse or current mode
X-ray transmission imaging Scintillation camera	Point source Collimation	Grid or air gap Pulse height discrimination	Current Pulse

raphy and fluoroscopy, an image is projected on the image receptor because the x-rays originate from a very small source that approximates a point. In comparison, the photons in nuclear imaging are emitted isotropically throughout the patient and therefore collimation is necessary to form a projection image. Furthermore, in x-ray transmission imaging, scattered x-rays can be distinguished from primary x-rays by their directions and thus can be largely removed by grids. In emission imaging, primary photons cannot be distinguished from scattered photons by direction. The collimator removes about the same fraction of scattered photons as it does primary photons and, unlike the grid in x-ray transmission imaging, does not reduce the fraction of counts in the resultant image due to scatter. Scattered photons in nuclear imaging can only be differentiated from primary photons by energy, because scattering reduces photon energy. In emission imaging, energy discrimination is used to reduce the fraction of counts in the image caused by scattered radiation. Finally, nuclear imaging devices must use pulse mode (the signal from each interaction is processed separately) for event localization and so that interaction-by-interaction energy discrimination can be employed; x-ray transmission imaging systems have photon fluxes that are too high to allow a radiation detector to detect, discriminate, and record on a photon-by-photon basis.

Performance

Measures of Performance

Measures of the performance of a scintillation camera with the collimator attached are called *system* or *extrinsic* measurements. Measures of camera performance with the collimator removed are called *intrinsic* measurements. System measurements give the best indication of clinical performance, but intrinsic measurements are often more useful for comparing the performance of different cameras, because they isolate camera performance from collimator performance.

Uniformity is a measure of a camera's response to uniform irradiation of the detector surface. The ideal response is a perfectly uniform image. Intrinsic uniformity is measured by placing a point radionuclide source (typically 150 μCi of Tc-99m) in front of the uncollimated camera. The source should be placed at least four times the largest dimension of the crystal away to ensure uniform irradiation of the camera surface and at least five times away, if the uniformity image is to be analyzed quantitatively using a computer. System uniformity, which reveals collimator as well as camera defects, is assessed by placing a uniform planar radionuclide source in

front of the collimated camera. Solid planar sealed sources of Co-57 (typically 5 to 10 mCi) and planar sources that may be filled with a Tc-99m solution are commercially available. A planar source should be large enough to cover the crystal of the camera. The uniformity of the resultant images may be analyzed by a computer or evaluated visually.

Spatial resolution is a measure of a camera's ability to accurately portray spatial variations in activity concentration and to distinguish as separate radioactive objects in close proximity. The *system spatial resolution* is evaluated by acquiring an image of a line source, such as a capillary tube filled with Tc-99m, using a computer interfaced to the collimated camera and determining the line spread function (LSF). The LSF, described in Chapter 10, is a cross-sectional profile of the image of a line source. The full-width-at-half-maximum (FWHM), the full-width-at-tenth-maximum (FWTM), and the modulation transfer function (MTF, described in Chapter 10) may all be derived from the LSF.

The system spatial resolution, if measured in air so that scatter is not present, is determined by the collimator resolution and the intrinsic resolution of the camera. The *collimator resolution* is defined as the FWHM of the radiation transmitted through the collimator from a line source (Fig. 21-8). It is not directly measured, but calculated from the system and intrinsic resolutions. The *intrinsic resolution* is determined quantitatively by acquiring an image with a sheet of lead containing thin slits placed against the uncollimated camera using a point source. The point source should be positioned at least five times the largest dimension of the camera's crystal away. The system (R_s), collimator (R_c), and intrinsic (R_i) resolutions, as indicated by the FWHMs of the LSFs, are related by the following equation:

$$R_s = \sqrt{R_c^2 + R_i^2} \quad [21-1]$$

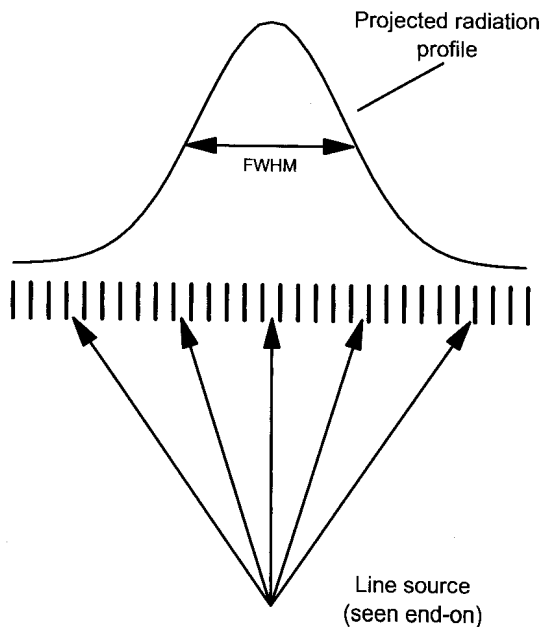


FIGURE 21-8. Collimator resolution.

Some types of collimators magnify (converging, pinhole) or minify (diverging) the image. For these collimators, the system and collimator resolutions should be corrected for the collimator magnification so they refer to distances in the object rather than distances in the camera crystal:

$$R'_S = \sqrt{R_C'^2 + (R_I/m)^2} \quad [21-2]$$

where $R'_S = R_S/m$ is the system resolution corrected for magnification, $R'_C = R_C/m$ is the collimator resolution corrected for magnification, and m is the collimator magnification (image size in crystal/object size). The collimator magnification is determined as follows:

- $m = 1.0$ for parallel-hole collimators,
- $m = (\text{pinhole-to-crystal distance})/(\text{object-to-pinhole distance})$ for pinhole collimators,
- $m = f/(f - x)$ for converging collimators, and
- $m = f/(f + x)$ for diverging collimators,

where f is the distance from the crystal to the focal point of the collimator, and x is the distance of the object from the crystal. It will be seen later in this section that the collimator and system resolutions degrade (FWHM of the LSF, corrected for collimator magnification, increases) with increasing distance between the line source and collimator.

As can be seen from Equation 21-2, collimator magnification reduces the deleterious effect of intrinsic spatial resolution on the overall system resolution. Geometric magnification in x-ray transmission imaging reduces the effect of image receptor blur for the same reason.

In routine practice, the intrinsic resolution is semiquantitatively evaluated by imaging a parallel line or four-quadrant bar phantom (Fig. 21-9) in contact with the camera face, using a planar radionuclide source if the camera is collimated or a

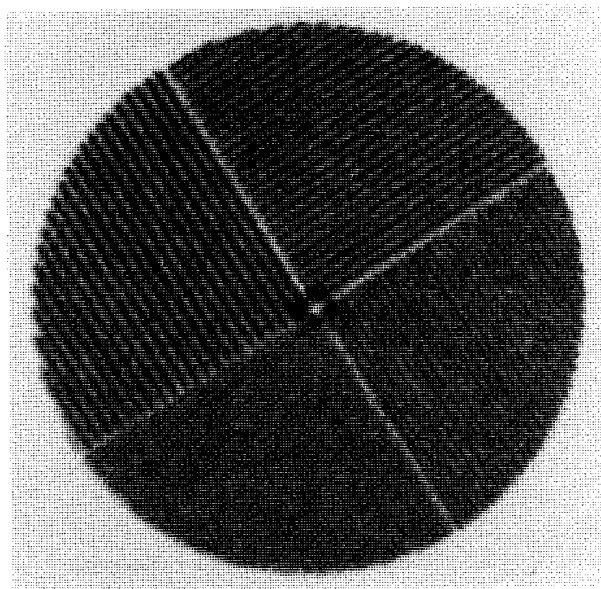


FIGURE 21-9. Image of bar phantom for evaluating spatial resolution. The lead bars and spaces between the bars are of equal widths. A typical four-quadrant bar phantom for a state-of-the-art scintillation camera might have 2.0, 2.5, 3.0, and 3.5 mm wide lead bars. The bar phantom is placed against the crystal of the uncollimated camera and irradiated by a point source at least four camera crystal diameters away.

distant point source if the collimator is removed, and visually noting the smallest bars that are resolved. The size of the smallest bars that are resolved is approximately related to the FWHM of the LSF:

$$(\text{FWHM of the LSF}) \approx 1.7 \times (\text{Size of smallest bars resolved}) \quad [21-3]$$

Spatial linearity (lack of spatial distortion) is a measure of the camera's ability to portray the shapes of objects accurately. It is determined from the images of a bar phantom, line source, or other phantom by assessing the straightness of the lines in the image. Spatial nonlinearities can significantly degrade the uniformity, as will be discussed later in this chapter.

Multienergy spatial registration, commonly called *multiple window spatial registration*, is a measure of the camera's ability to maintain the same image magnification, regardless of the energies deposited in the crystal by the incident x- or gamma rays. (Image magnification is defined as the distance between two points in the image divided by the actual distance between the two points in the object being imaged.) Higher energy x- and gamma-ray photons produce larger signals from the individual PMTs than do lower energy photons. The position circuit of the scintillation camera normalizes the X- and Y-position signals by the energy signal so that the position signals are independent of the deposited energy. If a radionuclide emitting useful photons of several energies, such as Ga-67 (93-, 185-, and 300-keV gamma rays), is imaged and the normalization is not properly performed, the resultant image will be a superposition of images of different magnifications (Fig. 21-10). The multienergy spatial registration can be tested by imaging several point sources of Ga-67, offset from the center of the camera, using only one major gamma ray at a time. The centroid of the count distribution of each source should be at the same position in the image for all three gamma ray energies.

The *system efficiency* (sensitivity) of a scintillation camera is the fraction of x- or gamma rays emitted by a source that produces counts in the image. The system efficiency is important because it, in conjunction with imaging time, determines the amount of quantum mottle (graininess) in the images. The system efficiency (E_s) is

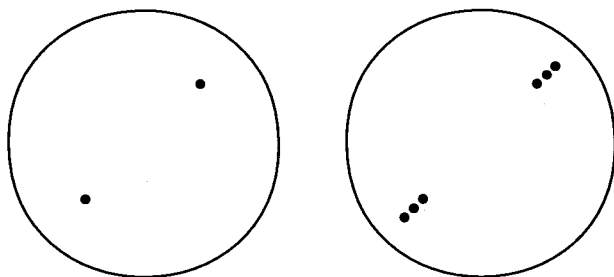


FIGURE 21-10. Multienergy spatial registration. **Left:** A simulated image of two point sources of a radionuclide emitting gamma rays of three different energies, illustrating proper normalization of the X- and Y-position signals for deposited energy. **Right:** A simulated image of the same two point sources, showing improper adjustment of the energy normalization circuit. The maladjustment causes higher energy photons to produce larger X and Y values than lower energy photons interacting at the same position in the camera's crystal, resulting in multiple images of each point source. A less severe maladjustment would cause each point source to appear as an ellipse, rather than three discrete dots.

the product of three factors: the collimator efficiency (E_c), the intrinsic efficiency of the crystal (E_i), and the fraction (f) of interacting photons accepted by the energy discrimination circuits:

$$E_s = E_c \times E_i \times f \quad [21-4]$$

The *collimator efficiency* is the fraction of photons emitted by a source that penetrate the collimator holes. In general, it is a function of the distance between the source and the collimator and the design of the collimator. The *intrinsic efficiency*, the fraction of photons penetrating the collimator that interact with the NaI(Tl) crystal, is determined by the thickness of the crystal and the energy of the photons:

$$E_i = 1 - e^{-\mu x} \quad [21-5]$$

where μ is the linear attenuation coefficient of sodium iodide and x is the thickness of the crystal. The last two factors in Equation 21-4 can be combined to form the *photopeak efficiency*, defined as the fraction of photons reaching the crystal that produce counts in the photopeak of the energy spectrum:

$$E_p = E_i \times f \quad [21-6]$$

This equation assumes that the window of the energy discrimination circuit has been adjusted to encompass the photopeak exactly. In theory, the system efficiency and each of its components range from zero to one. In reality, low-energy all-purpose (LEAP) parallel-hole collimators have efficiencies of about 2×10^{-4} and low-energy high-resolution parallel-hole collimators have efficiencies of about 1×10^{-4} .

The *energy resolution* of a scintillation camera is a measure of its ability to distinguish between interactions depositing different energies in its crystal. A camera with superior energy resolution is able to reject a larger fraction of photons that have scattered in the patient and coincident interactions, thereby producing images of better contrast. The energy resolution is measured by exposing the camera to a point source of a radionuclide, usually Tc-99m, emitting monoenergetic photons and acquiring a spectrum of the energy (Z) pulses, using either the nuclear medicine computer interfaced to the camera, if it has the capability, or a multichannel analyzer. The energy resolution is calculated from the full-width-at-half-maximum (FWHM) of the photopeak (see Pulse Height Spectroscopy in Chapter 20). The FWHM is divided by the energy of the photon (140 keV for Tc-99m) and is expressed as a percentage. A wider FWHM implies poorer energy resolution.

Scintillation cameras are operated in pulse mode (see Design and Principles of Operation, above) and therefore suffer from dead-time count losses at high interaction rates. They behave as paralyzable systems (see Chapter 20); the indicated count rate initially increases as the imaged activity increases, but ultimately reaches a maximum and decreases thereafter. The count-rate performance of a camera is usually specified by (a) the observed count rate at 20% count loss and (b) the maximal count rate. These count rates may be measured with or without scatter. Both are reduced when measured with scatter. Table 21-2 lists typical values for modern cameras. These high count rates are usually achieved by implementing a high count-rate mode that degrades the spatial and energy resolutions.

Some scintillation cameras with digital electronics can correctly process the PMT signals when two or more interactions occur simultaneously in the crystal, provided that the interactions are separated by sufficient distance. This significantly increases the number of interactions that are correctly registered at high interaction rates.

TABLE 21-2. TYPICAL INTRINSIC PERFORMANCE CHARACTERISTICS OF A MODERN SCINTILLATION CAMERA, MEASURED BY NEMA PROTOCOL^a

Intrinsic spatial resolution (FWHM of LSF for 140 keV)	2.7 to 4.2 mm
Energy resolution (FWHM of photopeak for 140 keV photons)	9.2% to 11%
Integral uniformity (max. pixel – min. pixel)/(max. pixel + min. pixel)	2% to 5%
Absolute spatial linearity	Less than 1.5 mm
Observed count rate at 20% count loss (measured without scatter)	110,000 to 260,000 counts/sec
Observed maximal count rate (measured without scatter)	170,000 to 500,000 counts/sec

^aIntrinsic spatial resolution is for a 0.95-cm (3/8-inch) thick crystal; thicker crystals cause slightly worse spatial resolution.

FWHM, full width at half maximum; LSF, line spread function; NEMA, National Electrical Manufacturers Association.

Design Factors Determining Performance

Intrinsic Spatial Resolution and Intrinsic Efficiency

The intrinsic spatial resolution of a scintillation camera is determined by the types of interactions by which the x- or gamma rays deposit energy in the camera's crystal and the statistics of the detection of the visible light photons emitted following these interactions. The most important of these is the random error associated with the collection of visible light photons and subsequent production of electrical signals by the PMTs. Approximately one visible light photon is emitted in the NaI crystal for every 25 eV deposited by an x- or gamma ray. For example, when a 140-keV gamma ray is absorbed by the crystal, approximately $140,000/25 = 5,600$ visible light photons are produced. About two-thirds of these, approximately 3,700 photons, reach the photocathodes of the PMTs. Only about one out of every five of these causes an electron to escape a photocathode, giving rise to about 750 electrons. These electrons are divided among the PMTs, with the most being produced in the PMTs closest to the interaction. Thus, only a small number of photoelectrons is generated in any one PMT after an interaction. Because the processes by which the absorption of a gamma ray causes the release of electrons from a photocathode are random, the pulses from the PMTs after an interaction contain significant random errors that, in turn, cause errors in the X and Y signals produced by the position circuit of the camera and the energy (Z) signal. These random errors limit both the intrinsic spatial resolution and the energy resolution of the camera. The energy of the incident x- or gamma rays determines the amount of random error in the event localization process; higher energy photons provide lower relative random errors and therefore superior intrinsic spatial resolution. For example, the gamma rays from Tc-99m (140 keV) produce better spatial resolution than do the x-rays from Tl-201 (69 to 80 keV), because each Tc-99m gamma ray produces about twice as many light photons when absorbed. There is relatively little improvement in intrinsic spatial resolution for gamma-ray energies above 250 keV because the improvement in spatial resolution due to more visible light photons is largely offset by the increased likelihood of scattering in the crystal before photoelectric absorption (discussed below).

The quantum detection efficiency of the PMTs in detecting the visible light photons produced in the crystal is the most significant factor limiting the intrinsic spatial resolution (typically only 20% to 25% of the visible light photons incident on a photo-

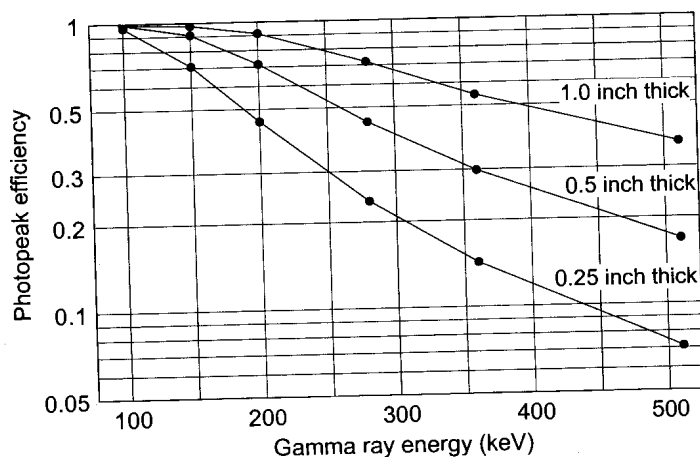


FIGURE 21-11. Calculated photopeak efficiency as a function of x- or gamma-ray energy for NaI(Tl) crystals. (Data from Anger HO, Davis DH. Gamma-ray detection efficiency and image resolution in sodium iodide. *Rev Sci Instr* 1964;35:693.)

cathode contribute to the signal from the PMT). The size of the PMTs also affects the spatial resolution; using PMTs of smaller diameter improves the spatial resolution by providing better sampling of the light emitted following each interaction in the crystal.

A thinner NaI crystal provides better intrinsic spatial resolution than a thicker crystal. A thinner crystal permits less spreading of the light before it reaches the PMTs. Furthermore, a thinner crystal reduces the likelihood of an incident x- or gamma ray undergoing one or more Compton scatters in the crystal followed by photoelectric absorption. Compton scattering in the crystal can cause the centroid of the energy deposition to be significantly offset from the site of the initial interaction in the crystal. The likelihood of one or more scatters preceding the photoelectric absorption increases with the energy of the x- or gamma ray.

The intrinsic efficiency of a scintillation camera is determined by the thickness of the crystal and the energy of the incident x- or gamma rays. Figure 21-11 is a graph of photopeak efficiency (fraction of incident x- or gamma rays producing counts in the photopeak) as a function of photon energy for NaI(Tl) crystals 0.635 cm ($\frac{1}{4}$ inch), 1.27 cm ($\frac{1}{2}$ inch), and 2.54 cm (1 inch) thick. Most modern cameras have 0.95 cm ($\frac{3}{8}$ inch) thick crystals. The photopeak efficiency of these crystals is greater than 80% for the 140-keV gamma rays from Tc-99m, but less than 30% for the 364-keV gamma rays from I-131. Some cameras, designed primarily for imaging radionuclides such as Tl-201 and Tc-99m, which emit low energy photons, have 0.635 cm ($\frac{1}{4}$ inch) thick crystals. With increasing interest in imaging the 511-keV annihilation photons of fluorine 18 fluorodeoxyglucose, some cameras have 1.27 to 2.5 cm ($\frac{1}{2}$ to 1 inch) thick crystals.

There is a design compromise between the intrinsic efficiency of the camera, which increases with crystal thickness, and intrinsic spatial resolution, which degrades with crystal thickness. This design compromise is similar to the design compromise between the spatial resolution of x-ray intensifying screens, which deteriorates with increasing screen thickness, and detection efficiency, which improves with increasing screen thickness.

Collimator Resolution and Collimator Efficiency

The collimator spatial resolution, as defined above, of multihole collimators (parallel-hole, converging, and diverging) is determined by the geometry of the holes. The

spatial resolution improves (narrower FWHM of the LSF) as the diameters of the holes are reduced and the lengths of the holes (thickness of the collimator) are increased. Unfortunately, changing the hole geometry to improve the spatial resolution in general reduces the collimator's efficiency. *The resultant compromise between collimator efficiency and collimator resolution is the single most significant limitation on scintillation camera performance.*

The spatial resolution of a parallel-hole collimator decreases (i.e., FWHM of the LSF increases) linearly as the collimator-to-object distance increases. This degradation of collimator spatial resolution with increasing collimator-to-object distance is also one of the most important factors limiting scintillation camera performance. Surprisingly, however, the efficiency of a parallel-hole collimator is nearly constant over the collimator-to-object distances used for clinical imaging. Although the number of photons passing through a particular collimator hole decreases as the square of the distance, the number of holes through which photons can pass increases as the square of the distance. The efficiency of a parallel-hole collimator, neglecting septal penetration, is

$$E_C = \frac{A}{4\pi l^2} g \quad [21-7]$$

where A is the cross-sectional area of a single collimator hole, l is the length of a hole (i.e., the thickness of the collimator), and g is the fraction of the frontal area of the collimator that is not blocked by the collimator septa (g = total area of holes in collimator face/area of collimator face).

Figure 21-12 depicts the line spread function (LSF) of a parallel-hole collimator as a function of source-to-collimator distance. The width of the LSF increases

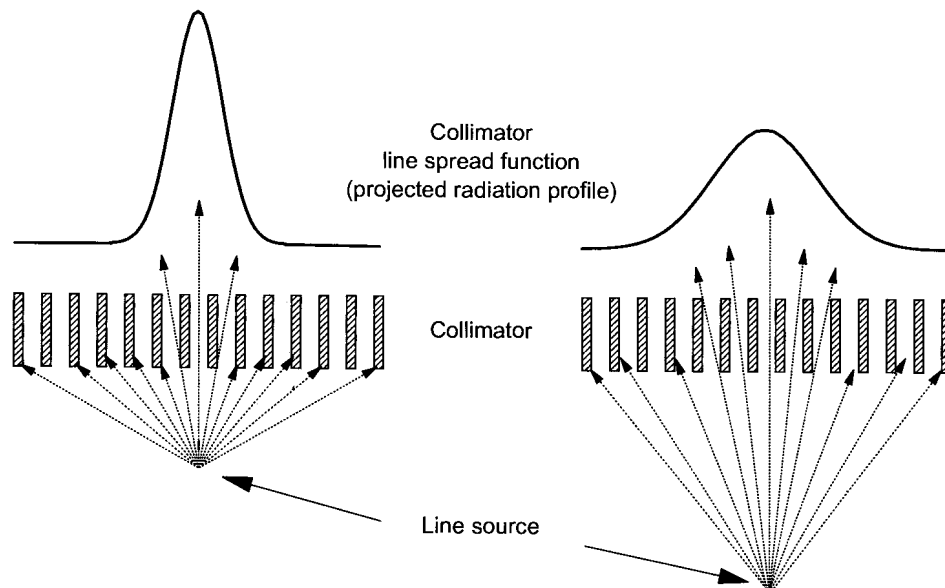


FIGURE 21-12. Line spread function (LSF) of a parallel-hole collimator as a function of source-to-collimator distance. The full-width-at-half-maximum (FWHM) of the LSF increases linearly with distance from the source to the collimator; however, the total area under the LSF (photon fluence through the collimator) decreases very little with source to collimator distance. (In both figures, the line source is seen "end-on.")

TABLE 21-3. THE EFFECT OF INCREASING COLLIMATOR-TO-OBJECT DISTANCE ON COLLIMATOR PERFORMANCE PARAMETERS

Collimator	Spatial resolution ^a	Efficiency	Field size	Magnification
Parallel hole	Decreases	Approximately constant	Constant	Constant ($m = 1.0$)
Converging	Decreases	Increases	Decreases	Increases ($m > 1$ at collimator surface)
Diverging	Decreases	Decreases	Increases	Decreases ($m < 1$ at collimator surface)
Pinhole	Decreases	Decreases	Increases	Decreases (m largest near pinhole)

^aSpatial resolution corrected for magnification.

with distance. Nevertheless, the area under the LSF (total number of counts) does not significantly decrease with distance.

The spatial resolution of a pinhole collimator, along its central axis, corrected for collimator magnification, is approximately equal to

$$R'_C \approx d \frac{f+x}{f} \quad [21-8]$$

where d is the diameter of the pinhole, f is the distance from the crystal to the pinhole, and x is the distance from the pinhole to the object. The efficiency of a pinhole collimator, along its central axis, neglecting penetration around the pinhole, is

$$E_C = d^2/16x^2 \quad [21-9]$$

Thus, the spatial resolution of the collimator improves (i.e., R'_C decreases) as the diameter of the pinhole is reduced, but the collimator efficiency decreases as the square of the pinhole diameter. Both the collimator spatial resolution and the efficiency are best for objects close to the pinhole (small x). The efficiency decreases rapidly with pinhole-to-object distance.

The spatial resolution of all collimators decreases with increasing gamma-ray energy because of increasing collimator septal penetration (or, in the case of pinhole collimators, penetration around the pinhole). Table 21-3 compares the characteristics of different types of collimators.

System Spatial Resolution and Efficiency

The system spatial resolution and efficiency are determined by the intrinsic and collimator resolutions and efficiencies, as described by Equations 21-2 and 21-4. Figure 21-13 (top) shows the effect of object-to-collimator distance on system spatial resolution. The system resolutions shown in this figure are corrected for magnification. *The system spatial resolution, when corrected for magnification, is degraded (i.e., FWHM of the LSF increases) as the collimator-to-object distance increases for all types of collimators.* This degradation of resolution with increasing patient-to-collimator distance is among the most important factors in image acquisition. Technologists should make every effort to minimize this distance during clinical imaging.

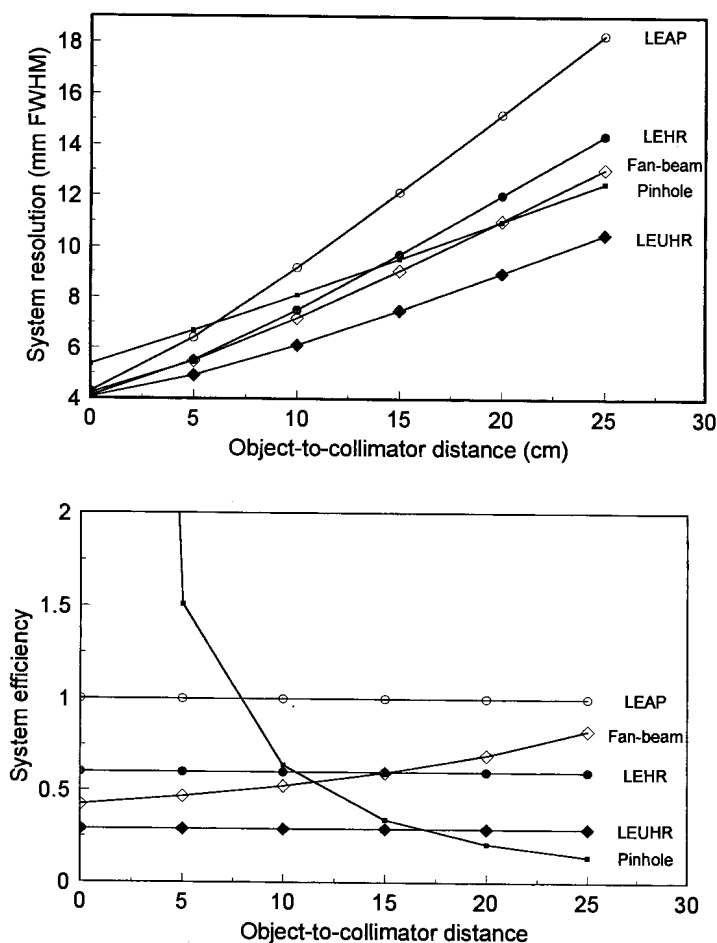


FIGURE 21-13. System spatial resolution (**top**) and efficiency (**bottom**) as a function of object-to-collimator distance (in cm). System resolutions for pinhole and fan-beam collimators are corrected for magnification. System efficiencies are relative to low-energy, all-purpose (LEAP) parallel-hole collimator (system efficiency 340 cpm/ μ Ci Tc-99m for a 0.95 cm thick crystal). LEHR, low-energy, high-resolution parallel-hole collimator; LEUHR, low-energy, ultra-high-resolution parallel-hole collimator. (Data courtesy of the late William Guth of Siemens Medical Systems, Nuclear Medicine Group.)

Figure 21-13 (bottom) shows the effect of object-to-collimator distance on system efficiency. The system efficiency with a parallel-hole collimator is nearly constant with distance. The system efficiency with a pinhole collimator decreases significantly with distance. The system efficiency of a fan-beam collimator (described in the following chapter) increases with collimator-to-object distance.

Modern scintillation cameras with 0.95 cm ($3/8$ inch) thick crystals typically have intrinsic spatial resolutions slightly less than 4.0 mm FWHM for the 140-keV gamma rays of Tc-99m. One manufacturer's low-energy high-resolution parallel-hole collimator has a resolution of 1.5 mm FWHM at the face of the collimator and

6.4 mm at 10 cm from the collimator. The system resolutions, calculated from Equation 21-1, are as follows:

$$R_S = \sqrt{(1.5 \text{ mm})^2 + (4.0 \text{ mm})^2} = 4.3 \text{ mm at } 0 \text{ cm}$$

and

$$R_S = \sqrt{(6.4 \text{ mm})^2 + (4.0 \text{ mm})^2} = 7.5 \text{ mm at } 10 \text{ cm}.$$

Thus, the system resolution is only slightly worse than the intrinsic resolution at the collimator face, but is largely determined by the collimator resolution at typical imaging distances for internal organs.

Spatial Linearity and Uniformity

Spatial nonlinearity (Fig. 21-14 top center and top right) is caused by the nonrandom (i.e., systematic) mispositioning of events. It is mainly due to interactions being shifted in the resultant image toward the center of the nearest PMT by the position circuit of the camera. Modern cameras have digital circuits that use tables of correction factors to correct each pair of X- and Y-position signals for spatial nonlinearities (Fig. 21-15). One lookup table contains an X-position correction for each portion of the crystal and another table contains corrections for the Y direction. Each pair of digital position signals is used to "look up" a pair of X and Y corrections in the tables. The corrections are added to the uncorrected X and Y values. The corrected X and Y values (X' and Y' in Fig. 21-15) may be sent directly to a

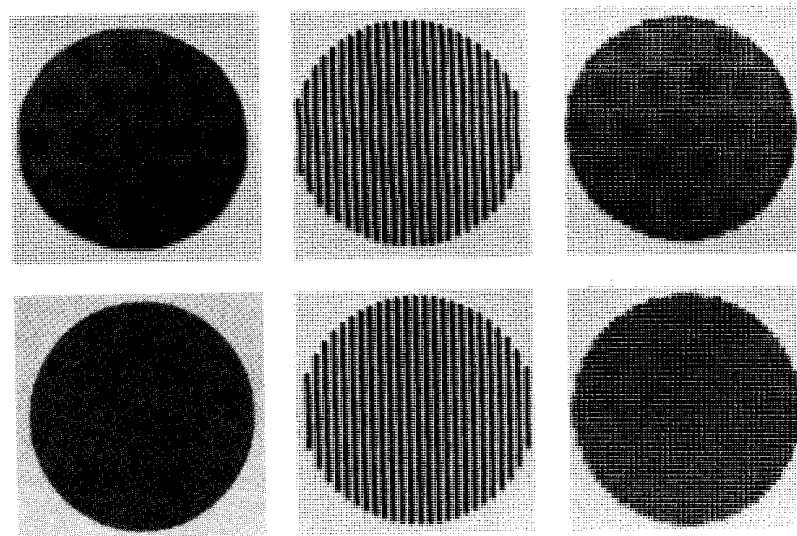


FIGURE 21-14. Pairs of uniformity images, lead slit-mask (lead sheet with thin slits) images, and orthogonal hole phantom (lead sheet with a rectangular array of holes) images, with scintillation camera's digital correction circuitry disabled (**top**) to demonstrate nonuniformities and spatial nonlinearities inherent to a scintillation camera and with correction circuitry functioning (**bottom**), demonstrating effectiveness of linearity and energy (Z) signal correction circuitry. (Photographs courtesy of Everett W. Stoub, Ph.D., formerly of Siemens Gammasonics, Inc.)

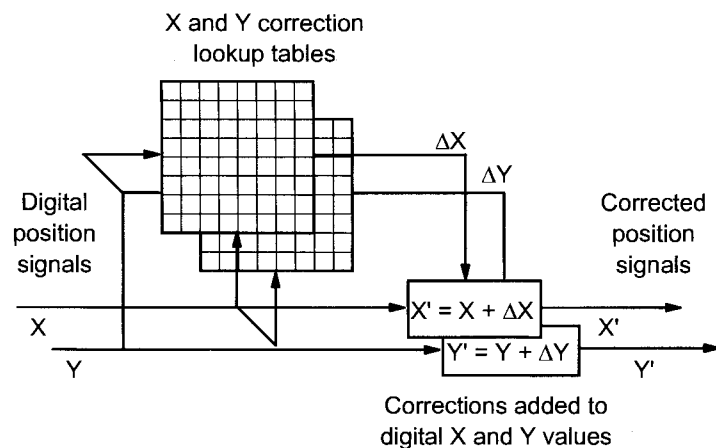


FIGURE 21-15. Spatial linearity correction circuitry of a scintillation camera.

computer or other device that accepts them in digital form or may be converted to analog form by a pair of DACs for devices that cannot accept them in digital form.

There are three major causes of nonuniformity. The first is spatial nonlinearities. As shown in Fig. 21-16, the systematic mispositioning of events imposes local variations in the count density. Spatial nonlinearities that are almost imperceptible can cause significant nonuniformities. The linearity correction circuitry previously described effectively corrects this source of nonuniformity.

The second major cause of nonuniformity is that the position of the interaction in the crystal affects the magnitude of the energy (Z) signal. It may be caused by local variations in the crystal in the light generation and light transmission to the PMTs and by variations in the light detection and gains of the PMTs. If these

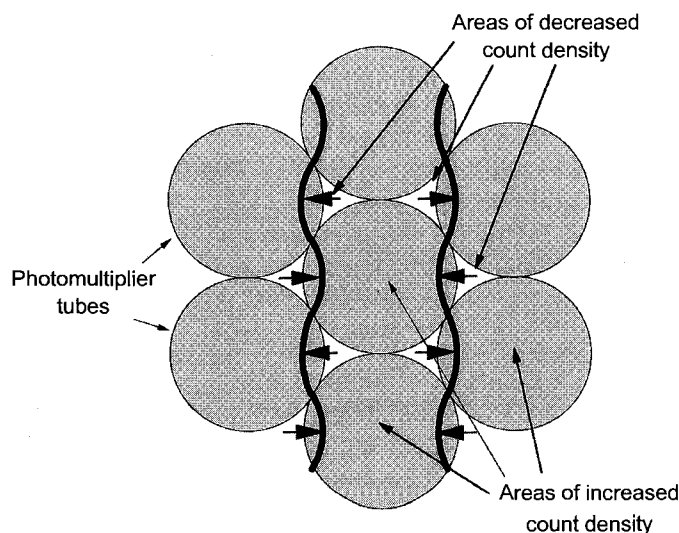


FIGURE 21-16. Spatial nonlinearities cause nonuniformities. The two vertical wavy lines represent straight lines in the object that have been distorted. The scintillation camera's position circuit causes the locations of individual counts to be shifted toward the center of the nearest photomultiplier tube (PMT), causing an enhanced count density toward the center of the PMT and a decreased count density between the PMTs, as seen in the top left image in Fig. 21-12.

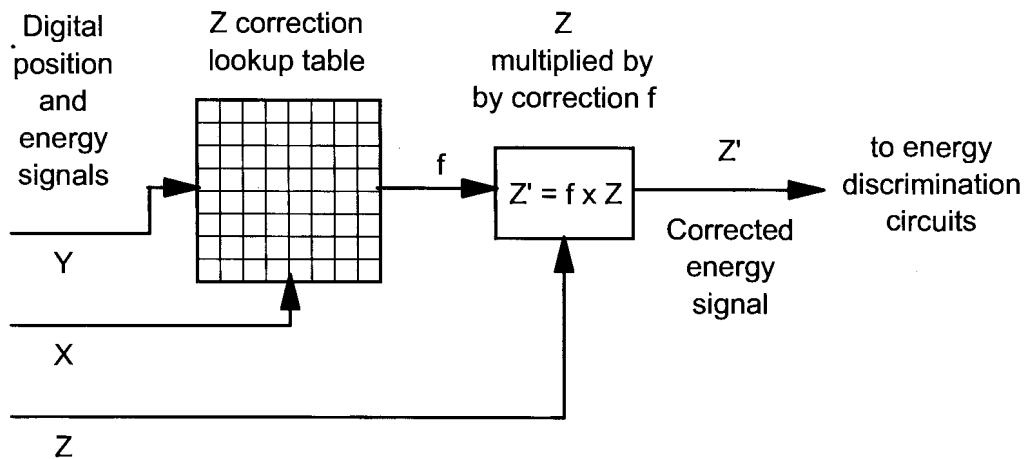


FIGURE 21-17. Energy (Z) signal correction circuitry of a scintillation camera. The digital X- and Y-position signals are used to “look up” a correction factor. The uncorrected Z value is multiplied by the correction factor. Finally, the corrected Z value is transmitted to the energy discrimination circuits.

regional variations in energy signal are not corrected, the fraction of interactions rejected by the energy discrimination circuits will vary with position in the crystal. These positional variations in the magnitude of the energy signal are corrected by digital electronic circuitry in modern cameras (Fig. 21-17).

The third major cause of nonuniformity is local variations in the efficiency of the camera in absorbing x- or gamma rays, such as manufacturing defects or damage to the collimator. These may be corrected by acquiring an image of an extremely uniform planar source using a computer interfaced to the camera. A correction factor is determined for each pixel in the image by dividing the average pixel count by that pixel count. Each pixel in a clinical image is multiplied by its correction factor to compensate for this cause of nonuniformity.

The digital spatial linearity and energy correction circuits of a modern scintillation camera do not directly affect its intrinsic spatial resolution. However, the obsolete fully analog cameras, which lacked these corrections, used means such as thick light pipes and light absorbers between the crystal and PMTs to provide adequate linearity and uniformity. These reduced the amount of light reaching the PMTs. As mentioned previously, the intrinsic spatial resolution and the energy resolution are largely determined by the statistics of the detection of light photons from the crystal. The loss of signal from light photons in fully analog cameras significantly limited their intrinsic spatial and energy resolutions. The use of digital linearity and energy correction permits camera designs that maximize light collection, thereby providing much better intrinsic spatial resolution and energy resolution than fully analog cameras.

NEMA Specifications for Performance Measurements of Scintillation Cameras

The National Electrical Manufacturers Association (NEMA) publishes a document entitled *Performance Measurements of Scintillation Cameras* that specifies standard

methods for measuring the camera performance. All manufacturers of scintillation cameras publish NEMA performance measurements of their cameras, which are used by prospective purchasers in comparing cameras and for writing purchase specifications. Before the advent of NEMA standards, manufacturers measured the performance of scintillation cameras in a variety of ways, making it difficult to compare the manufacturers' specifications for different cameras objectively. The measurement methods specified by NEMA require specialized equipment and are not intended to be performed by the nuclear medicine department. However, it is possible to perform simplified versions of the NEMA tests to determine whether a newly purchased camera meets the published specifications. Unfortunately, the NEMA protocols omit testing of a number of important camera performance parameters.

Effects of Scatter and Attenuation on Projection Images

Ideally, a nuclear medicine projection image would be a two-dimensional projection of the three-dimensional activity distribution in the patient. If this were the case, the number of counts in each point in the image would be proportional to the average activity concentration along a straight line through the corresponding anatomy of the patient. There are three main reasons why nuclear medicine images are not ideal projection images—attenuation of photons in the patient, inclusion of Compton scattered photons in the image, and the degradation of spatial resolution with distance from the collimator.

Attenuation in the patient by Compton scattering and the photoelectric effect prevents some photons that would otherwise pass through the collimator holes from contributing to the image. The amount of attenuation is mainly determined by the path length through tissue and the densities of the tissues between a location in the patient and the corresponding point on the camera face. Thus, photons from structures deeper in the patient are much more heavily attenuated than photons from structures closer to the camera face. Attenuation is more severe for lower energy photons, such as the 68- to 80-keV characteristic x-rays emitted by Tl-201 ($\mu \approx 0.19 \text{ cm}^{-1}$), than for higher energy photons, such as the 140-keV gamma rays from Tc-99m ($\mu = 0.15 \text{ cm}^{-1}$). Nonuniform attenuation, especially in thoracic and cardiac nuclear imaging, presents a particular problem in image interpretation.

The vast majority of the interactions with soft tissue of x- and gamma rays of the energies used for nuclear medicine imaging are by Compton scattering. Some photons that have scattered in the patient pass through the collimator holes and are detected. As in diagnostic radiology, the relative number of scattered photons is greater when imaging thicker parts of the patient, such as the abdomen, and the main effect of counts in the image from scattered photons is a loss of contrast. As was mentioned earlier in this chapter, the number of scattered photons contributing to the image is reduced by pulse height discrimination. However, setting an energy window of sufficient width to encompass most of the photopeak permits a considerable amount of the scatter to contribute to image formation.

Operation and Routine Quality Control

Peaking a scintillation camera means to adjust its energy discrimination windows to center them on the photopeak or photopeaks of the desired radionuclide. There are

two ways that scintillation cameras commonly display pulse-height spectra—the Z-pulse display (on older cameras) and the multichannel analyzer (MCA) type display (Fig. 21-18). The Z-pulse display shows a superposition of the actual Z (energy) pulses; photopeaks appear as especially bright bands on the display. On this type of display, the energy window is portrayed as a dark rectangle. The midpoint energy (E) of the window is the distance from the bottom of the display to the center of the dark rectangle and the window width (ΔE) is the vertical dimension of the rec-

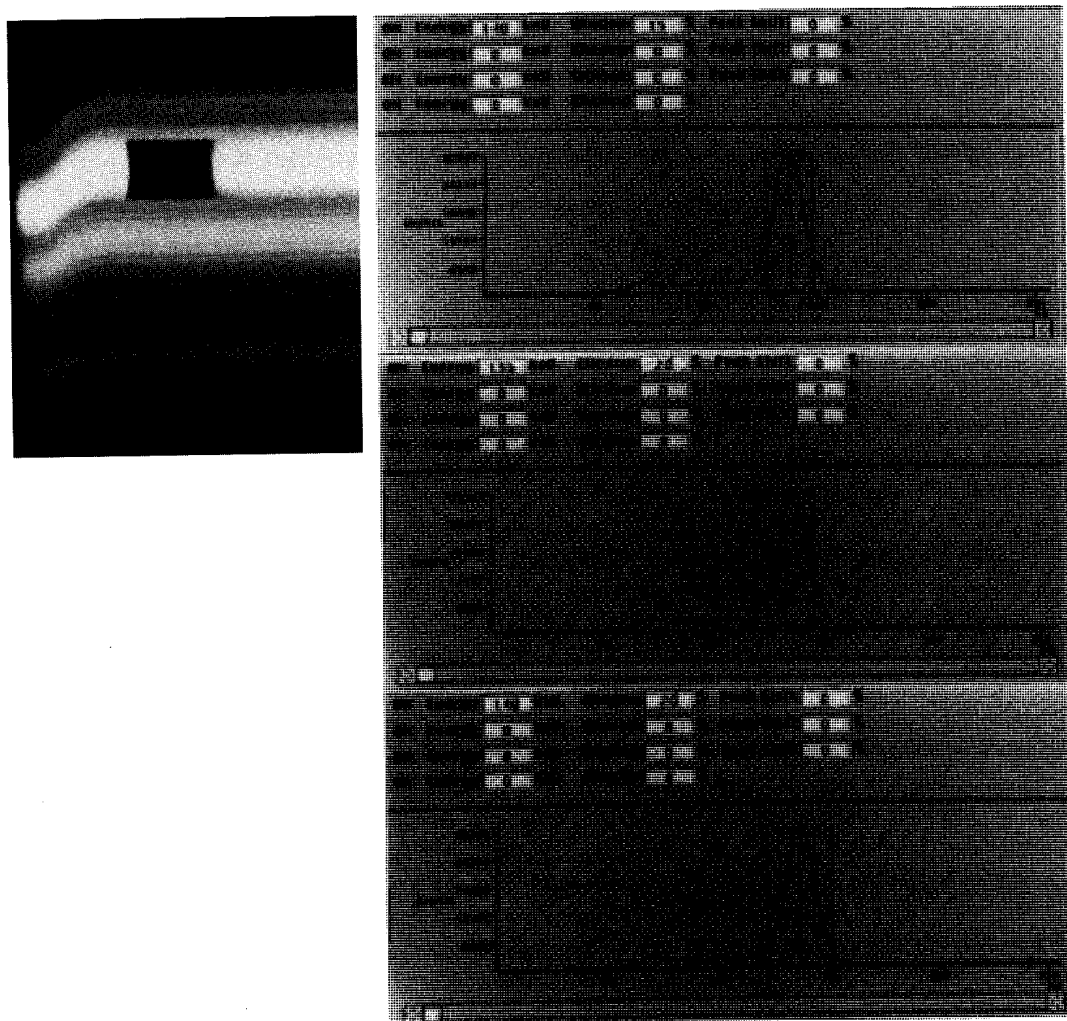


FIGURE 21-18. Z pulse (*left*) and MCA type (*right*) displays of a pulse-height spectrum used for “peaking” the scintillation camera [centering the single-channel analyzer (SCA) window on the photopeak]. In the older Z-pulse display, photon energy is the vertical scale and brightness indicates the rate of interactions. The horizontal bright band is the photopeak. The dark rectangle represents the SCA window. Adjusting the energy (E) setting moves the rectangle up and down; adjusting the window width (ΔE) increases the vertical size of the rectangle. When properly adjusted, the rectangle is centered on the bright band. In the more modern multichannel analyzer (MCA) type display, the energy window is shown by the two vertical lines around the photopeak. **Top:** A properly adjusted 15% window. **Middle:** An improperly adjusted 20% window. **Bottom:** A properly adjusted 20% window. (Photographs courtesy of Siemens Medical Systems, Nuclear Medicine Group.)

tangle. The MCA-type display presents the spectrum as a histogram of the number of interactions as a function of pulse height, like the display produced by a multi-channel analyzer (see Chapter 20). On this display, the energy window limits are shown as vertical lines. A narrower energy window provides better scatter rejection, but also reduces the number of unscattered photons recorded in the image.

Peaking may be performed manually by adjusting the energy window settings while viewing the spectrum or automatically by the camera. If a camera is peaked automatically, the technologist should view the spectral display to ensure that the energy window is centered upon the photopeak. A scintillation camera should be peaked before first use each day and before imaging a different radionuclide. A small source should be used to peak the camera; using the radiation emitted by the patient to peak a camera would constitute poor practice because of its large scatter component.

The uniformity of the camera should be assessed daily and after each repair. The assessment may be made intrinsically by using a Tc-99m point source, or the system uniformity may be evaluated by using a Co-57 planar source or a fillable flood source. Uniformity images must contain sufficient counts so that quantum mottle does not mask uniformity defects. About 2 million counts should be obtained for the daily test of a standard FOV camera [30-cm (12-inch) diameter FOV], about 3.5 million counts should be obtained for the daily test of a large FOV camera [41-cm (16-inch) diameter FOV], and as many as 6 million counts for the largest rectangular head cameras. Uniformity images can be evaluated visually or can be analyzed by a computer program, avoiding the subjectivity of visual evaluation. The uniformity test will reveal most malfunctions of a scintillation camera. Figure 21-19 shows uniformity images from a modern scintillation camera. Multi-hole collimators are easily damaged by careless handling, often by striking them with an imaging table. Technologists should inspect the collimators on each camera daily and whenever changing collimators. Old damage should be marked. The uniformity of each collimator should be evaluated periodically, by using a Co-57 planar source or a fillable flood source and whenever damage is suspected. The fre-

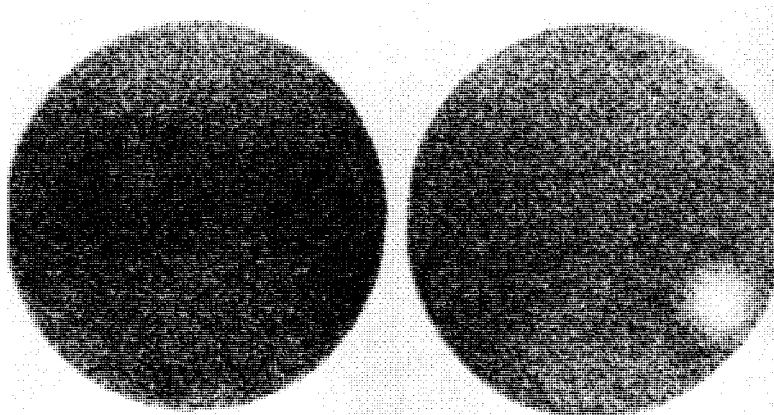


FIGURE 21-19. Uniformity images. **Left:** Image from a modern high-performance camera with digital spatial linearity and Z pulse correction circuitry. **Right:** The same camera, with an inoperative photomultiplier tube.

quency of this testing depends on the care taken by the technologists to avoid damaging the collimators and their reliability in inspecting the collimators and reporting new damage.

The spatial resolution and spatial linearity should be assessed at least weekly. If a four-quadrant bar phantom (Fig. 21-9) is used, it should be imaged four times with a 90-degree rotation between images to ensure all quadrants of the camera are evaluated. If a parallel-line phantom is used, it should be imaged twice with a 90-degree rotation between images.

The efficiency of each camera head should be measured periodically. It can be monitored during uniformity tests by always performing the test with the source at the same distance, determining the activity of the source, recording the number of counts and counting time, and calculating the count rate per unit activity.

Each camera should also have a complete evaluation at least annually, which includes not only the items listed above, but also multienergy spatial registration and count-rate performance. Table 22-1 in Chapter 22 contains a recommended schedule for quality control testing of scintillation cameras. Relevant tests should also be performed after each repair or adjustment of a scintillation camera.

New cameras should receive acceptance testing by an experienced medical physicist to determine if the camera's performance meets the purchase specifications and to establish baseline values for routine quality control testing.

Computers, Whole Body Scanning, and SPECT

Almost all scintillation cameras are connected to digital computers. The computer is used for the acquisition, processing, and display of digital images and for control of the movement of the camera heads. Computers used with scintillation cameras are discussed later in this chapter.

Many large-FOV scintillation cameras can perform whole-body scanning. Some systems move the camera past a stationary patient, whereas others move the patient table past a stationary scintillation camera. Moving camera systems require less floor space. In either case, the system must sense the position of the camera relative to the patient table and must add values that specify the position of the camera relative to the table to the X- and Y-position signals. Systems with round or hexagonal crystals usually must make two passes over each side of the patient. Newer large-area rectangular-head cameras can scan each side of all but extremely obese patients with a single pass, saving considerable time and producing superior image statistics.

Many modern computer-equipped scintillation cameras have heads that can rotate automatically around the patient and acquire images from different views. The computer mathematically manipulates these images to produce cross-sectional images of the activity distribution in the patient. This is called single photon emission computed tomography (SPECT) and is discussed in detail in Chapter 22.

Obtaining High-Quality Images

Attention must be paid to many technical factors to obtain high-quality images. Imaging procedures should be optimized for each type of study. Sufficient counts must be acquired so that quantum mottle in the image does not mask lesions. Imaging times must be as long as possible consistent with patient throughput and lack of patient motion. The camera or table scan speed should be sufficiently slow during whole-body scans to obtain adequate image statistics. The use of a higher reso-

lution collimator may improve spatial resolution and the use of a narrower energy window to reduce scatter may improve image contrast.

Because the spatial resolution of a scintillation camera is degraded significantly as the collimator-to-patient distance is increased, the camera heads should always be as close to the patient as possible. In particular, thick pillows and mattresses on imaging tables should be discouraged when images are acquired with the camera head below the table.

Significant nonuniformities can be caused by careless treatment of the collimators. Collimators are easily damaged by collisions with imaging tables or by placing them on top of other objects when they are removed from the camera.

Patient motion and metal objects worn by patients or in the patients' clothing are common sources of artifacts. Every effort should be made to identify metal objects and to remove them from the patient or from the field of view. The technologist should remain in the room during imaging to minimize patient motion. Furthermore, for safety, a patient should never be left unattended under a moving camera, either in whole-body or SPECT mode.

The results from routine technical quality control (QC) testing must be reviewed, not just performed. A QC test should not be considered completed until it has been reviewed, results recorded or plotted, and necessary corrective action taken or scheduled. Also, a deficiency revealed by a QC test should cause a review of any previous clinical images whose results may have been compromised.

21.2 COMPUTERS IN NUCLEAR IMAGING

Most nuclear medicine computer systems consist of commercially available computers with additional components to enable them to acquire, process, and display images from scintillation cameras. Figure 21-20 shows a typical system. Some man-



FIGURE 21-20. Modern nuclear medicine computer system. (Courtesy of Siemens Medical Systems, Nuclear Medicine Group.)

ufacturers incorporate a computer for image acquisition and camera control in the camera itself and provide a separate computer for image processing and display, whereas others provide a single computer for both purposes.

Some older scintillation cameras transmit pairs of analog X- and Y-position pulses to the computer. A computer interfaced to such a camera has an acquisition interface containing two ADCs for converting the pairs of analog position pulses into digital form. However, in most modern scintillation cameras, the pairs of X and Y signals are created using digital circuitry and are transferred from the camera to the computer in digital form. The formation of a digital image is described in the following section.

Digital Image Formats in Nuclear Medicine

As discussed in Chapter 4, a digital image consists of a rectangular array of numbers, and an element of the image represented by a single number is called a *pixel*. In nuclear medicine, each pixel represents the number of counts detected from activity in a specific location within the patient. Common image formats in nuclear medicine are 64^2 and 128^2 pixels. Whole-body images are stored in larger formats, such as 256 by 1,024 pixels. If one byte (8 bits) is used for each pixel, the image is said to be in *byte mode*; if two bytes are used for each pixel, the image is said to be in *word mode*. A single pixel can store as many as 255 counts in byte mode and 65,535 counts in word mode. Modern nuclear medicine computers may allow only word-mode images.

Image Acquisition

Frame Mode Acquisition

Image data in nuclear medicine are acquired in either frame or list mode. Figure 21-21 illustrates frame-mode acquisition. Before acquisition begins, a portion of the computer's memory is designated to contain the image. All pixels within this image are set to zero. After acquisition begins, pairs of X- and Y-position signals are received from the camera. Each pair of numbers designates a single pixel in the image. One count is added to the counts in that pixel. As many pairs of position signals are received, the image is formed.

There are three types of frame-mode acquisition: *static*, *dynamic*, and *gated*. In a static acquisition, a single image is acquired for either a preset time interval or until the total number of counts in the image reaches a preset number. In a dynamic acquisition, a series of images is acquired one after another, for a preset time per image. Dynamic image acquisition is used to study dynamic processes, such as the first transit of a bolus of a radiopharmaceutical through the heart or the extraction and excretion of a radiopharmaceutical by the kidneys.

Some dynamic processes occur too rapidly for them to be effectively portrayed by dynamic image acquisition; each image in the sequence would have too few counts to be statistically valid. However, if the dynamic process is repetitive, gated image acquisition may permit the acquisition of an image sequence that accurately depicts the dynamic process. Gated acquisition is most commonly used for evalu-

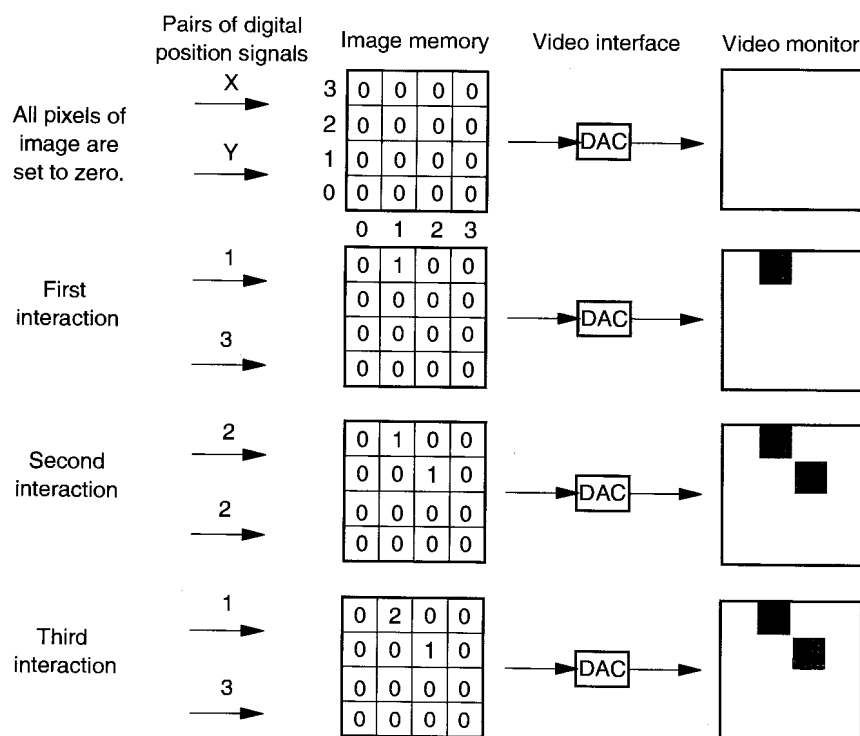


FIGURE 21-21. Acquisition of an image in frame mode. This figure illustrates the incorporation of position data from the first three interactions into an image.

ating cardiac mechanical performance, in cardiac blood pool studies and, sometimes, in myocardial perfusion studies. Gated acquisition requires a physiologic monitor that provides a trigger pulse to the computer at the beginning of each cycle of the process being studied. In gated cardiac studies, an electrocardiogram (ECG) monitor provides a trigger pulse to the computer whenever the monitor detects a QRS complex.

Figure 21-22 depicts the acquisition of a gated cardiac image sequence. First, space for the desired number of images (usually 16 to 24) is reserved in the computer's memory. Next, several cardiac cycles are timed and the average time per cycle is determined. The time per cycle is divided by the number of images in the sequence to obtain the time T per image. Then the acquisition begins. When the first trigger pulse is received, the acquisition interface sends the data to the first image in the sequence for a time T . Then it is stored in the second image in the sequence for a time T , after which it is stored in the third image for a time T . This process proceeds until the next trigger pulse is received. Then the process begins anew, with the data being added to that in the first image for a time T , then the second image for a time T , etc. This process is continued until a preset time interval, typically 10 minutes, has elapsed, enabling sufficient counts to be collected for the image sequence to form a statistically valid depiction of an average cardiac cycle.

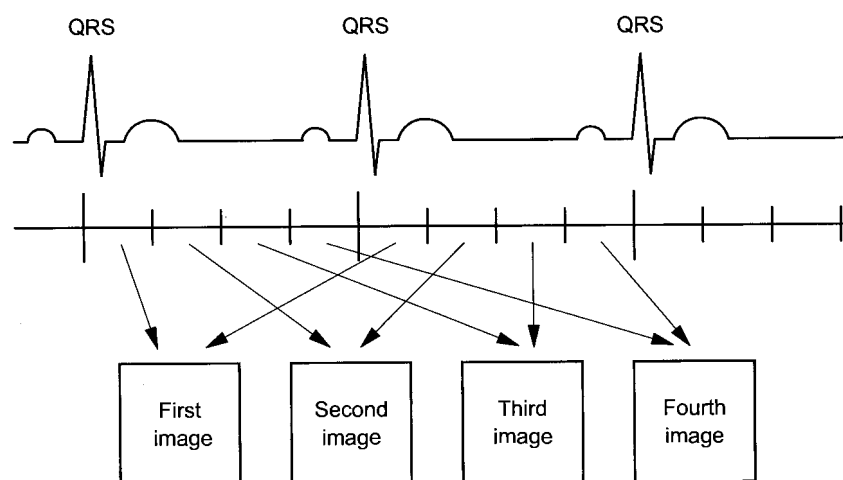


FIGURE 21-22. Acquisition of a gated cardiac image sequence. Only four images are shown here. Sixteen to 24 images are typically acquired.

List-Mode Acquisition

In *list-mode acquisition*, the pairs of X- and Y-position values are stored in a list (Fig. 21-23, instead of being immediately formed into an image. Periodic timing marks are included in the list. If a physiologic monitor is being used, as in gated cardiac imaging, trigger marks are also included in the list. After acquisition is complete, the list-mode data are reformatted into conventional images for display. The advantage of list-mode acquisition is that it allows great flexibility in how the X and Y values are combined to form images. The disadvantages of list-mode acquisition are that it generates large amounts of data, requiring more memory to acquire and disk space to store than frame-mode images, and that the data must be subsequently processed into standard images for viewing.

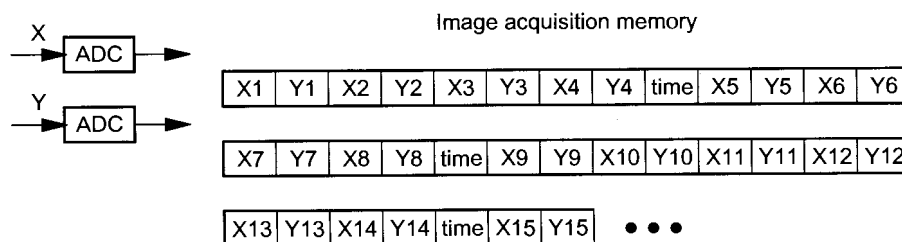


FIGURE 21-23. Acquisition of image data in list mode. The digital X- and Y-position signals are stored as a list. Periodically, a timing signal is also stored.

Image Processing in Nuclear Medicine

A major reason for the use of computers in nuclear medicine is that they provide the ability to present the data in the unprocessed images in ways that are of greater use to the clinician. Although it is not within the scope of this section to provide a comprehensive survey of image processing in nuclear medicine, the following are common examples.

Image Subtraction

When one image is subtracted from another, each pixel count in one image is subtracted from the corresponding pixel count in the other. Negative numbers resulting from these subtractions are set to zero. The resultant image depicts the change in activity that occurs in the patient during the time interval between the acquisitions of the two images.

Regions of Interest and Time-Activity Curves

A *region of interest* (ROI) is a closed boundary that is superimposed on the image. It may be drawn manually or it may be drawn automatically by the computer. The sum of the counts in all pixels in the ROI is an indication of the activity in the corresponding portion of the patient.

To create a *time-activity curve* (TAC), a ROI must first be drawn on one image of a dynamic or gated image sequence. The same ROI is then superimposed on each image in the sequence by the computer and the total number of counts within the ROI is determined for each image. Finally, the counts within the ROI are plotted as a function of image number. The resultant curve is an indication of the activity in the corresponding portion of the patient as a function of time. Figure 21-24 shows ROIs over both kidneys of a patient and time-activity curves describing the uptake and excretion of the radiopharmaceutical Tc-99m MAG-3 by the kidneys.

Spatial Filtering

Nuclear medicine images have a grainy appearance because of the statistical nature of the acquisition process. This quantum mottle can be reduced by a type of spatial filtering called *smoothing*. Unfortunately, smoothing also reduces the spatial resolution of the image. Images should not be smoothed to the extent that clinically significant detail is lost. Spatial filtering is described in detail in Chapter 11.

Left Ventricular Ejection Fraction

The left ventricular ejection fraction (LVEF) is a measure of the mechanical performance of the left ventricle of the heart. It is defined as the fraction of the end-diastolic volume ejected during a cardiac cycle:

$$\text{LVEF} = (V_{\text{ED}} - V_{\text{ES}})/V_{\text{ED}} \quad [21-10]$$

where V_{ED} is the end-diastolic volume and V_{ES} is the end-systolic volume of the ventricle. In nuclear medicine, it is usually determined from an equilibrium gated blood-

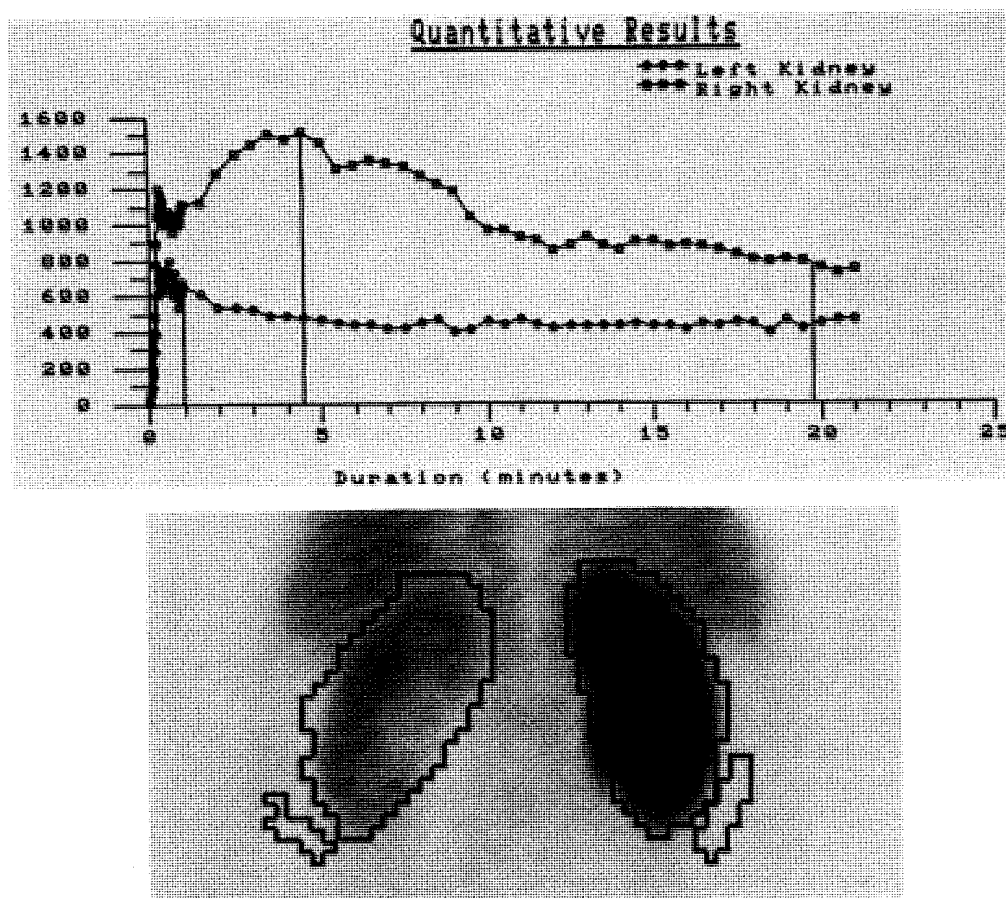


FIGURE 21-24. Regions of interest (ROIs) (**bottom**) and time-activity curves (TACs) (**top**).

pool image sequence, using Tc-99m-labeled red blood cells. The image sequence is acquired from a left anterior oblique (LAO) projection, with the camera positioned at the angle demonstrating the best separation of the two ventricles, after sufficient time has elapsed for the administered activity to reach a uniform concentration in the blood.

The calculation of the LVEF is based on the assumption that the counts from left ventricular activity are approximately proportional to the ventricular volume throughout the cardiac cycle. A ROI is first drawn around the left ventricular cavity, and a TAC is obtained by superimposing this ROI over all images in the sequence. The first image in the sequence depicts end diastole, and the image containing the least counts in the ROI depicts end systole. The total left ventricular counts in the end-diastolic and end-systolic images are determined. Some programs use the same ROI around the left ventricle for both images, whereas, in other programs, the ROI is drawn separately in each image to better fit the varying shape of the ventricle. Unfortunately, each of these counts is due not only to activity in the left ventricular cavity but also to activity in surrounding tissues, chambers, and great vessels. To compensate for this "crosstalk" (commonly called "background activity"), another ROI is drawn just beyond the wall of the left ventricle, avoiding active

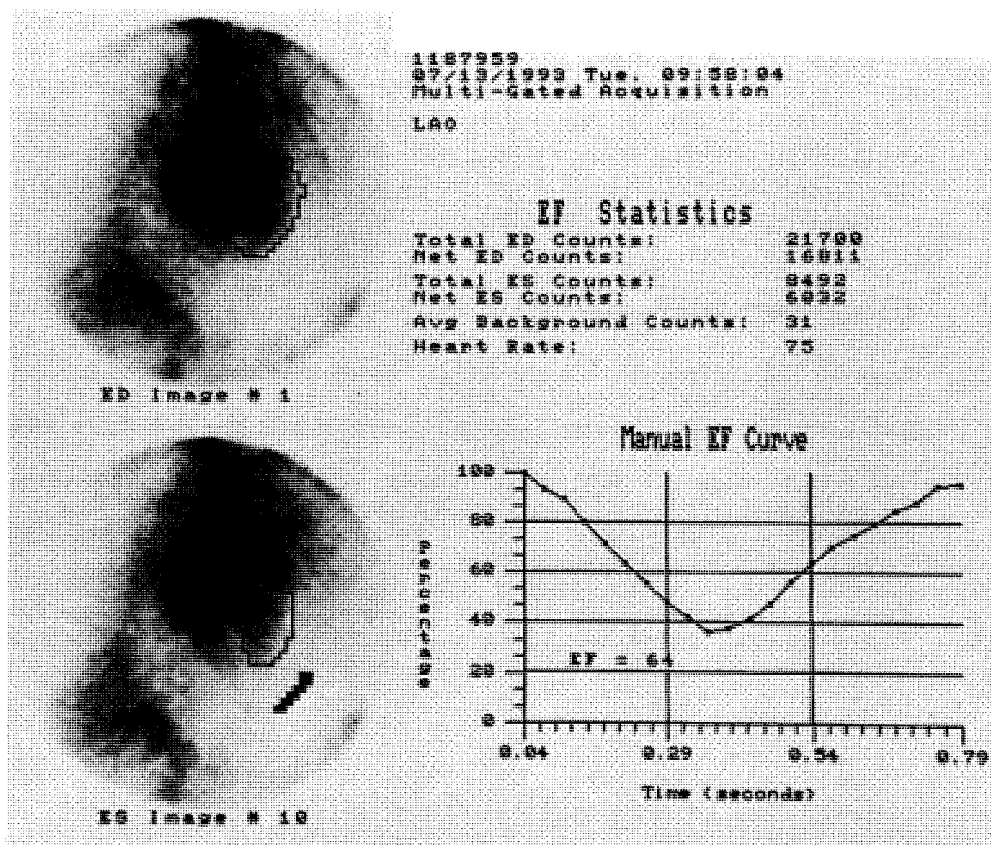


FIGURE 21-25. End-diastolic (**top**) and end-systolic (**bottom**) images from a gated cardiac blood-pool study, showing ROIs used to determine the left ventricular ejection fraction (LVEF). The small dark ROI below and to the right of the LV ROI in the end-systolic image is used to determine and correct for the extracardiac ("background") counts.

structures such as the spleen, cardiac chambers, and great vessels. The number of counts in the left-ventricular ROI due to crosstalk is estimated as follows:

$$\text{Counts crosstalk} = \frac{(\text{Counts in crosstalk ROI}) (\text{Pixels in LV ROI})}{(\text{Pixels in crosstalk ROI})} \quad [21-11]$$

Figure 21-25 shows end-diastolic and end-systolic images of the heart from the LAO projection and the ROIs used to determine the LVEF. The LVEF is then calculated using the following equation:

$$\text{LVEF} = \frac{(\text{Counts ED}) - (\text{Counts ES})}{(\text{Counts ED}) - (\text{Counts crosstalk})} \quad [21-12]$$

SUMMARY

This chapter described the principles of operation of the Anger scintillation camera, which forms images, depicting the distribution of x- and gamma-ray-emitting

radionuclides in patients, using a planar crystal of the scintillator NaI(Tl) optically coupled to a two-dimensional array of photomultiplier tubes. The collimator, necessary for formation of projection images, imposes a compromise between spatial resolution and sensitivity in detecting x- and gamma rays. Because of the collimator, the spatial resolution of the images degrades with distance from the face of the camera. In Chapter 22, the use of the scintillation camera to perform computed tomography, called *SPECT*, is described.

SUGGESTED READING

- Anger HO. Radioisotope cameras. In: Hine GJ, ed. *Instrumentation in nuclear medicine*, vol. 1. New York: Academic Press, 1967:485–552.
- Gelfand MJ, Thomas SR. *Effective use of computers in nuclear medicine*. New York: McGraw-Hill, 1988.
- Graham LS, Muehllehner G. Anger scintillation camera. In: Sandler MP, et al., eds. *Diagnostic nuclear medicine*, 3rd ed., vol. 1. Baltimore: Williams & Wilkins, 1996:81–92.
- Groch MW. Cardiac function: gated cardiac blood pool and first pass imaging. In: Henkin RE, et al., eds. *Nuclear medicine*, vol. 1. St. Louis: Mosby, 1996:626–643.
- Simmons GH, ed. *The scintillation camera*. New York: Society of Nuclear Medicine, 1988.
- Yester MV, Graham LS, eds. *Advances in nuclear medicine: the medical physicist's perspective*. Proceedings of the 1998 Nuclear Medicine Mini Summer School, American Association of Physicists in Medicine, June 21–23, 1998, Madison, WI.

NUCLEAR IMAGING—EMISSION TOMOGRAPHY

The formation of projection images in nuclear medicine was discussed in Chapter 21. A nuclear medicine projection image depicts a two-dimensional projection of the three-dimensional activity distribution in the patient. The disadvantage of a projection image is that the contributions to the image from structures at different depths overlap, hindering the ability to discern the image of a structure at a particular depth. Tomographic imaging is fundamentally different; it attempts to depict the activity distribution in a single cross section of the patient.

There are two fundamentally different types of tomography: conventional tomography, also called geometric or focal plane tomography, and computed tomography. In conventional tomography, structures out of a focal plane are not removed from the resultant image; instead, they are blurred by an amount proportional to their distance from the focal plane. Those close to the focal plane suffer little blurring and remain apparent in the image. Even those farther away, although significantly blurred, contribute to the image, thereby reducing contrast. In distinction, computed tomography (CT) uses mathematical methods to remove overlying structures completely. CT requires the acquisition of a set of projection images from at least a 180-degree arc about the patient. The projection images are then mathematically reconstructed by a computer to form images depicting cross sections of the patient. Just as in x-ray transmission imaging, both conventional and computed tomography are possible in nuclear medicine imaging. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) are both forms of CT.

Focal Plane Tomography in Nuclear Medicine

Focal plane tomography once had a significant role in nuclear medicine but is seldom used today. The rectilinear scanner, when used with focused collimators, was an example of conventional tomography. A number of other devices have been developed to exploit conventional tomography in nuclear medicine. The Anger tomoscanner used two small scintillation cameras with converging collimators, one above and one below the patient table, to scan the patient in a raster pattern and produce multiple whole-body images, each showing structures at a different depth in the patient in focus, in a single scan. The seven-pinhole collimator was used with a conventional scintillation camera and computer to produce short-axis images of the heart, each showing structures at a different depth in focus. The rectilinear scanner and the Anger tomoscanner are no longer produced. The seven-pinhole colli-

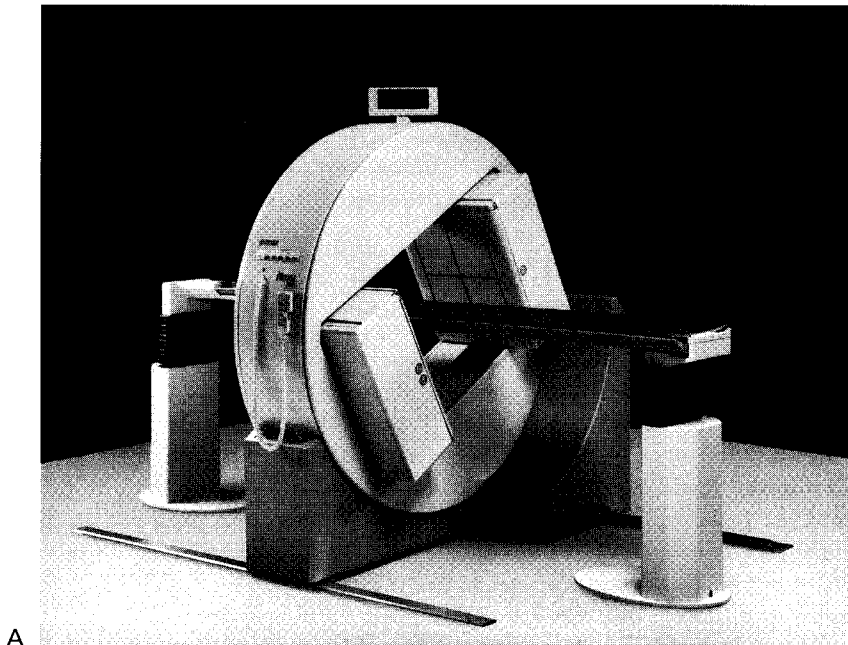
mator, which never enjoyed wide acceptance, has been almost entirely displaced by SPECT.

The scintillation camera itself, when used for planar imaging with a parallel-hole collimator, produces a weak tomographic effect. The system spatial resolution decreases with distance, causing structures farther from the camera to be more blurred than closer structures. This effect is perhaps most clearly evident in planar skeletal imaging of the body. In the anterior images, for example, the sternum and anterior portions of the ribs are clearly shown, while the spine and posterior ribs are barely evident. (Attenuation also plays a role in this effect.)

22.1 SINGLE PHOTON EMISSION COMPUTED TOMOGRAPHY (SPECT)

Design and Principles of Operation

Single photon emission computed tomography (SPECT) generates transverse images depicting the distribution of x- or gamma-ray emitting nuclides in patients. Standard planar projection images are acquired from an arc of 180 degrees (most cardiac SPECT) or 360 degrees (most noncardiac SPECT) about the patient. Although these images could be obtained by any collimated imaging device, the vast majority of SPECT systems use one or more scintillation camera heads that revolve about the patient. The SPECT system's digital computer then reconstructs the transverse images, using either filtered backprojection, as does the computer in an x-ray CT system, or iterative reconstruction methods, which are described later in this chapter. Figure 22-1 shows multihead SPECT systems. Figure 21-1 in Chapter 21 shows a single-head SPECT system.



A

FIGURE 22-1. A: Double-head, fixed 180-degree (photograph courtesy of Marconi Medical Systems).

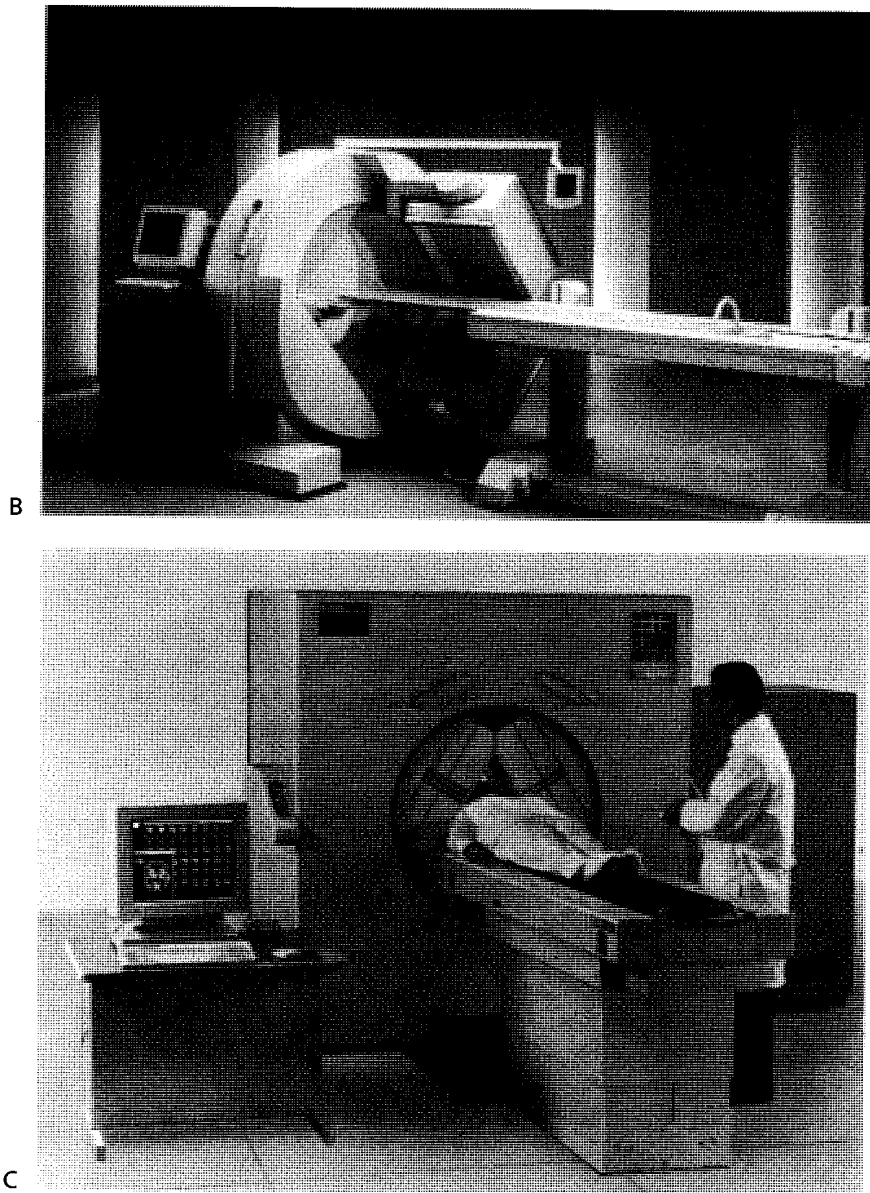


FIGURE 22-1 (continued). (B) double-head, variable-angle (photograph courtesy of Siemens Medical Systems); (C) and triple-head (photograph courtesy of Trionix Research Laboratory, Inc.) single photon emission computed tomography (SPECT) systems. The double-head, variable-angle system is shown with the heads at a 76-degree angle for cardiac acquisition.

Image Acquisition

The camera head or heads of a SPECT system revolve about the patient, acquiring projection images from evenly spaced angles. The head or heads may acquire the images while moving (continuous acquisition) or may stop at predefined angles to acquire the images ("step and shoot" acquisition). Each projection image is acquired in a computer in frame mode (see Image Acquisition in Chapter 21). If the camera

heads of a SPECT system produced ideal projection images (i.e., no attenuation by the patient and no degradation of spatial resolution with distance from the camera), projection images from opposite sides of the patient would be mirror images and projection images over a 180-degree arc would be sufficient for transverse image reconstruction. However, in SPECT, attenuation greatly reduces the number of photons from activity in the half of the patient opposite the camera head, and this information is greatly blurred by the distance from the collimator. Therefore, for most noncardiac studies, such as brain SPECT, the projection images are acquired over a complete revolution (360 degrees) about the patient. However, most nuclear medicine laboratories acquire cardiac SPECT studies, such as myocardial perfusion studies, over a 180-degree arc from the 45-degree right anterior oblique (RAO) view to the 45-degree left posterior oblique (LPO) view (Fig. 22-2). The 180-degree acquisition produces reconstructed images of superior contrast and resolution, because the projection images of the heart from the opposite 180 degrees have poor spatial resolution and contrast due to greater distance and attenuation. Although studies have shown that the 180-degree acquisition can introduce artifacts, the 180-degree acquisition is more commonly used than the 360-degree for cardiac studies.

SPECT projection images are usually acquired in either a 64^2 or a 128^2 pixel format. Using too small a pixel format reduces the spatial resolution of the projection images and of the resultant reconstructed transverse images. When the 64^2 format is used, typically 60 or 64 projection images are acquired and, when a 128^2 format is chosen, 120 or 128 projection images are acquired. Using too few projections creates radial streak artifacts in the reconstructed transverse images.

The camera heads on older SPECT systems followed circular orbits around the patient while acquiring images. Circular orbits are satisfactory for SPECT imaging of the brain, but cause a loss of spatial resolution in body imaging because the circular orbit causes the camera head to be many centimeters away from the surface of the body during the anterior and posterior portions of its orbit (Fig. 22-3). Newer SPECT systems provide noncircular orbits (also called "body contouring") that keep the camera heads in close proximity to the surface of the body throughout the orbit. For some systems, the technologist specifies the noncircular orbit by placing the camera head as close as possible to the patient at several angles, from which the camera's computer determines the orbit. Other systems perform automatic body

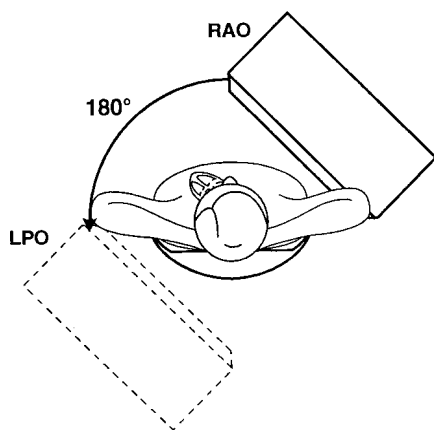


FIGURE 22-2. An 180-degree cardiac orbit.

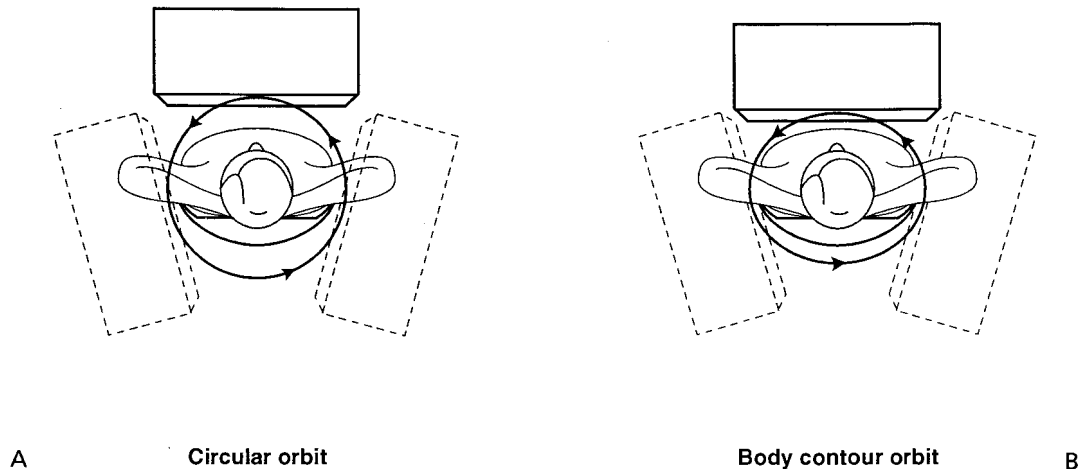


FIGURE 22-3. Circular (A) and body-contour (B) orbits.

contouring, using sensors on the camera heads to determine their proximity to the patient.

In brain SPECT, it is usually possible for the camera head to orbit with a much smaller radius than in body SPECT, thereby producing images of much higher spatial resolution. Unfortunately, in older cameras, the large distance from the physical edge of the camera head to the useful portion of the detector often made it impossible to orbit at a radius within the patient's shoulders while including the base of the brain in the images. These older systems were therefore forced to image the brain with an orbit outside the patient's shoulders, causing a significant loss of resolution. Most modern SPECT systems permit brain imaging with orbits within the patient's shoulders.

Transverse Image Reconstruction

After the projection images are acquired, they are usually corrected for axis-of-rotation misalignments and for nonuniformities. (These corrections are discussed below; see Quality Control in SPECT.) Following these corrections, transverse image reconstruction is performed using either filtered backprojection or iterative methods.

As described in Chapter 13, filtered backprojection consists of two steps. First, the projection images are mathematically filtered. Then, to form a particular transverse image, simple backprojection is performed of the row, corresponding to that transverse image, of each projection image. For example, the fifth row of each projection image is backprojected to form the fifth transverse image. A SPECT study produces transverse images covering the entire field of view (FOV) of the camera in the axial direction from each revolution of the camera head or heads.

Mathematical theory specifies that the ideal filter kernel, when displayed in frequency space, is the ramp filter (Fig. 22-4). However, the actual projection images contain considerable statistical noise. If they were filtered using a ramp filter kernel and then backprojected, the resultant transverse images would contain an unacceptable amount of statistical noise.

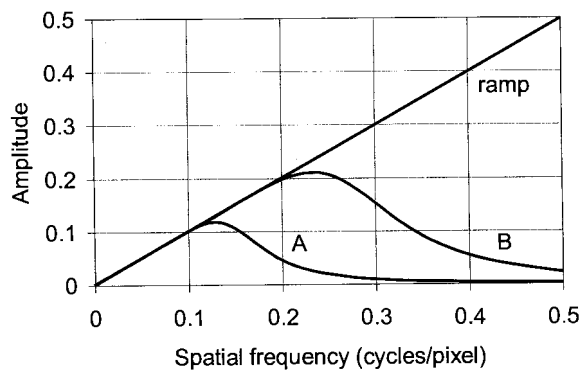


FIGURE 22-4. Typical filter kernels used for filtered backprojection. The kernels are shown in frequency space. Filter kernel A is a Butterworth filter of the 5th order with a critical frequency of 0.15 Nyquist, and filter kernel B is a Butterworth filter of the 5th order with a critical frequency of 0.27 Nyquist. Filter kernel A provides more smoothing than filter kernel B. A ramp filter, which provides no smoothing, is also shown.

In the spatial frequency domain, statistical noise predominates in the high-frequency portion. To smooth the projection images before backprojection, the ramp filter kernel is modified to “roll off” at higher spatial frequencies. Unfortunately, this reduces the spatial resolution of the projection images and thus of the reconstructed transverse images. A compromise must therefore be made between spatial resolution and statistical noise of the transverse images.

Typically, a different filter kernel is selected for each type of SPECT study; e.g., a different kernel would be used for HMPAO brain SPECT than would be used for Tl 201 myocardial perfusion SPECT. The choice of filter kernel for a particular type of study is determined by the amount of statistical noise in the projection images (mainly determined by the injected activity, collimator, and acquisition time per image) and their spatial resolution [determined by the collimator and the typical distances of the camera head(s) from the organ being imaged]. The preference of the interpreting physician regarding the appearance of the images also plays a role. Projection images of better spatial resolution and less quantum mottle require a filter with a higher spatial frequency cutoff to avoid unnecessary loss of spatial resolution in the reconstructed transverse images, whereas projection images of poorer spatial resolution and greater quantum mottle require a filter with a lower spatial frequency cutoff to avoid excessive quantum mottle in the reconstructed transverse images. Although the SPECT camera’s manufacturer may suggest filters for specific imaging procedures, the filters are usually empirically optimized in each nuclear medicine laboratory. Figure 22-5 shows a SPECT image created using three different filter kernels, illustrating too much smoothing, proper smoothing, and no smoothing.

Filtered backprojection is computationally efficient. However, it is based on the assumption that the projection images are perfect projections of a three-dimensional object. As discussed in the previous chapter, this is far from true in nuclear medicine imaging, mainly because of attenuation of photons in the patient, inclusion of Compton scattered photons in the image, and the degradation of spatial resolution with distance from the collimator.

In SPECT, iterative reconstruction methods are increasingly being used instead of filtered backprojection. In iterative methods, an initial activity distribution in the patient is assumed. Then, projection images are calculated from the assumed activity distribution, using the known imaging characteristics of the scintillation camera.

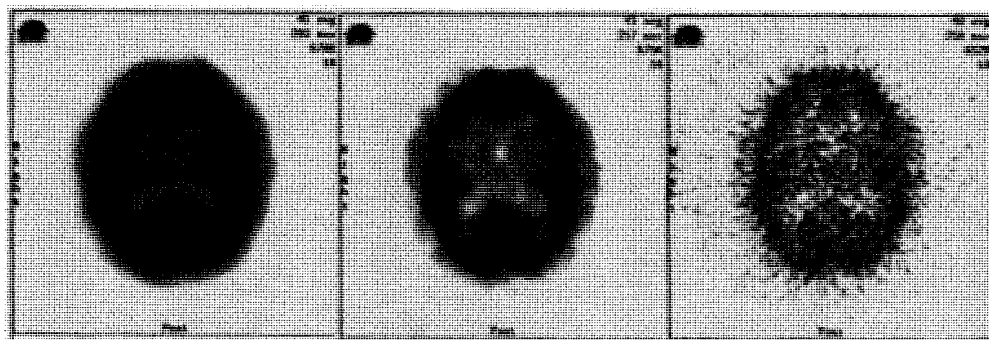


FIGURE 22-5. SPECT images created by filtered backprojection. The projection images were filtered using the filter kernels shown in Fig. 22-4. **Left:** An image produced using filter kernel A, which exhibits a significant loss of spatial resolution. **Center:** An image produced using filter kernel B, which provides a proper amount of smoothing. **Right:** An image produced using the ramp filter, which shows good spatial resolution but excessive statistical noise.

The calculated projection images are compared with the actual projection images and, based on this comparison, the assumed activity distribution is adjusted. This process is repeated several times, with successive adjustments to the assumed activity distribution, until the calculated projection images approximate the actual projection images.

As was stated above, in each iteration projection images are calculated from the assumed activity distribution. The calculation of projection images uses the point spread function of the scintillation camera, which takes into account the decreasing spatial resolution with distance from the camera face. The point spread function can be modified to incorporate the effect of photon scattering in the patient. Furthermore, if a map of the attenuation characteristics of the patient is available, the calculation of the projection images can include the effects of attenuation. If this is done, iterative methods will partially compensate for the effects of decreasing spatial resolution with distance, photon scattering in the patient, and attenuation in the patient.

Iterative methods are computationally less efficient than filtered backprojection. However, the increasing speed of computers, the small image matrix sizes used in nuclear imaging, and development of computationally efficient algorithms have made iterative reconstruction feasible for SPECT.

Attenuation Correction in SPECT

X- or gamma rays that must traverse long paths through the patient produce fewer counts, due to attenuation, than do those from activity closer to the near surface of the patient. For this reason, transverse image slices of a phantom with a uniform activity distribution, such as a jug filled with a well-mixed solution of radionuclide, will show a gradual decrease in activity toward the center (Fig. 22-6 left). Attenuation effects are more severe in body SPECT than in brain SPECT.

Approximate methods are available for attenuation correction. One of the most common, the Chang method, assumes a constant attenuation coefficient throughout the patient. Approximate attenuation corrections can over- or undercompensate

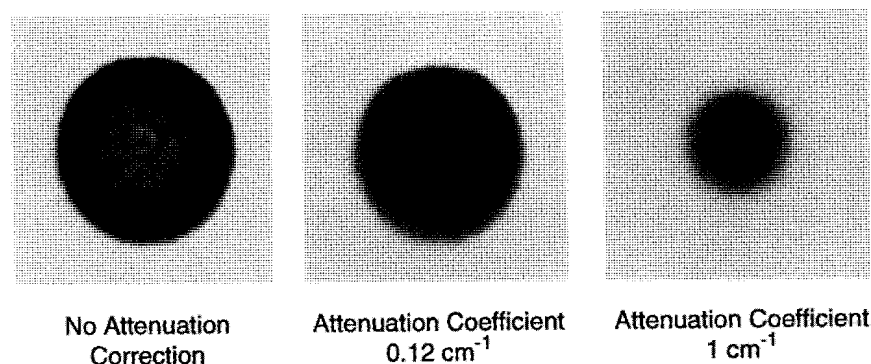


FIGURE 22-6. Attenuation correction. **Left:** A reconstructed transverse image slice of a cylindrical phantom containing a well-mixed radionuclide solution. This image shows a decrease in activity toward the center due to attenuation. (A small ring artifact, unrelated to the attenuation, is also visible in the center of the image.) **Center:** The same image corrected by the Chang method, using a linear attenuation coefficient of 0.12 cm^{-1} , demonstrating proper attenuation correction. **Right:** The same image, corrected by the Chang method using an excessively large attenuation coefficient.

for attenuation. If such a method is to be used, its proper functioning should be verified using phantoms before it is used in clinical studies.

Attenuation is not uniform throughout the patient and, in the thorax in particular, it is very nonuniform. These approximate methods cannot compensate for nonuniform attenuation.

Several manufacturers provide SPECT cameras with radioactive sources to measure the attenuation through the patient. The sources are used to acquire transmission data from projections around the patient. After acquisition, the transmission projection data are reconstructed to provide maps of tissue attenuation characteristics across transverse sections of the patient, similar to x-ray CT images. Finally, these attenuation maps are used during the SPECT image reconstruction process to provide attenuation-corrected SPECT images.

The transmission sources are available in several configurations. These include scanning collimated line sources that are used with parallel-hole collimators, arrays of fixed line sources used with parallel-hole collimators, and a fixed line source located at the focal point of a fan-beam collimator.

The transmission data are usually acquired simultaneously with the acquisition of the emission projection data, because performing the two separately poses significant problems in the spatial alignment of the two data sets. The radionuclide used for the transmission measurements is chosen to have primary gamma-ray emissions that differ significantly in energy from those of the radiopharmaceutical. Separate energy windows are used to differentiate the photons emitted by the transmission source from those emitted by the radiopharmaceutical. However, scattering of the higher energy photons in the patient and in the detector causes some crosstalk in the lower energy window.

In SPECT, attenuation correction using transmission sources has most often been applied in myocardial perfusion imaging, where attenuation artifacts can mimic perfusion defects. To date, however, clinical studies evaluating attenuation correction using transmission sources have not been uniformly favorable. At this

time, attenuation correction in SPECT using transmission sources is promising, but it is still under development and its ultimate clinical utility is not yet proven.

Generation of Coronal, Sagittal, and Oblique Images

The pixels from the transverse image slices may be reordered to produce coronal and sagittal images. For cardiac imaging, it is desirable to produce oblique images oriented either parallel (vertical and horizontal long axis images) or perpendicular (short axis images) to the long axis of the left ventricle. Because there is considerable anatomic variation among patients regarding the orientation of the long axis of the left ventricle, the long axis of the heart must be marked on the displayed image.

SPECT Collimators

The most commonly used collimator for SPECT is the high-resolution parallel-hole collimator. However, specialized collimators have been developed for SPECT. The fan-beam collimator, shown in Fig. 22-7, is a hybrid of the converging and the parallel-hole collimators. Because it is a parallel-hole collimator in the y-direction, each row of pixels in a projection image corresponds to a single transaxial slice of the subject. In the x-direction, it is a converging collimator, with spatial resolution-efficiency characteristics that are superior to those of a parallel-hole collimator (see Chapter 21, Fig. 21-13). Because a fan-beam collimator is a converging collimator in the cross-axial direction, its FOV decreases with distance from the collimator. For this reason, the fan-beam collimator is mainly used for brain SPECT; if the collimator is used for body SPECT, portions of the body are excluded from the FOV, creating artifacts in the reconstructed images.

Multihead SPECT Cameras

To reduce the limitations imposed on SPECT by collimation and limited time per view, camera manufacturers provide SPECT systems with two or three scintillation

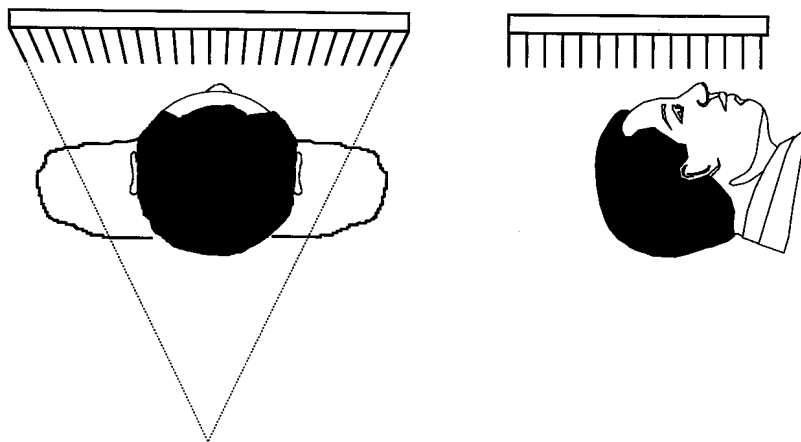


FIGURE 22-7. Fan-beam collimator.

camera heads that revolve about the patient (Fig. 22-1). The use of multiple camera heads permits the use of higher resolution collimators, for a given level of quantum mottle in the images, than would a single head system. However, the use of multiple camera heads imposes considerable technical difficulties on the manufacturer. It places severe requirements on the electrical and mechanical stability of the camera heads. Furthermore, the Y-offsets and X- and Y-magnification factors of all the heads must be precisely matched throughout rotation. Early multihead systems failed to achieve acceptance because of the difficulty of matching these factors. Today's multihead systems are more mechanically and electronically stable than earlier systems, providing high-quality tomographic images for a variety of clinical applications.

Multihead scintillation cameras are available in several configurations. Double-head cameras fixed in a 180-degree configuration are good for head and body SPECT and whole-body planar scans. Triple-head, fixed-angle cameras are good for head and body SPECT, but less suitable for whole-body planar scans because of the limited widths of the crystals. Double-head, variable-angle cameras are highly versatile, capable of head and body SPECT and whole-body planar scans in the 180-degree configuration and cardiac SPECT in the 90-degree configuration. (The useful portion of the NaI crystal does not extend all the way to the edge of a camera head. If the two camera heads are placed at an angle of exactly 90 degrees to each other, both heads cannot be close to the patient without parts of the patient being outside of the fields of view. For this reason, one manufacturer provides SPECT acquisitions with the heads at a 76-degree angle to each other.)

Performance

Spatial Resolution

The spatial resolution of a SPECT system can be measured by acquiring a SPECT study of a line source, such as a capillary tube filled with a solution of technetium (Tc) 99m, placed parallel to the axis of rotation. The National Electrical Manufacturers Association (NEMA) specifies a cylindrical plastic water-filled phantom, 22 cm in diameter, containing three line sources (Fig. 22-8, left) for measuring spatial resolution. The full-width-at-half-maximum (FWHM) measures of the line sources are determined from the reconstructed transverse images (Fig. 22-8, right). A ramp filter is used in the filtered backprojection, so that the filtering does not reduce the spatial resolution. The NEMA spatial resolution measurements are primarily determined by the collimator used. The tangential resolution for the peripheral sources (typically 7 to 8 mm FWHM for low-energy high-resolution and ultra-high-resolution parallel-hole collimators) is much superior to the central resolution (typically 9.5 to 12 mm). The tangential resolution for the peripheral sources is better than the radial resolution (typically 9.4 to 12 mm) for the peripheral sources.

These FWHMs measured using the NEMA protocol, while providing a useful index of ultimate system performance, are not necessarily representative of clinical performance, because these spatial resolution studies can be acquired using longer imaging times and closer orbits than would be possible in a patient. Patient studies may require the use of lower resolution (higher efficiency) collimators than the one used in the NEMA measurement to obtain adequate image statistics. In addition, the filters used before backprojection for clinical studies have lower spatial fre-

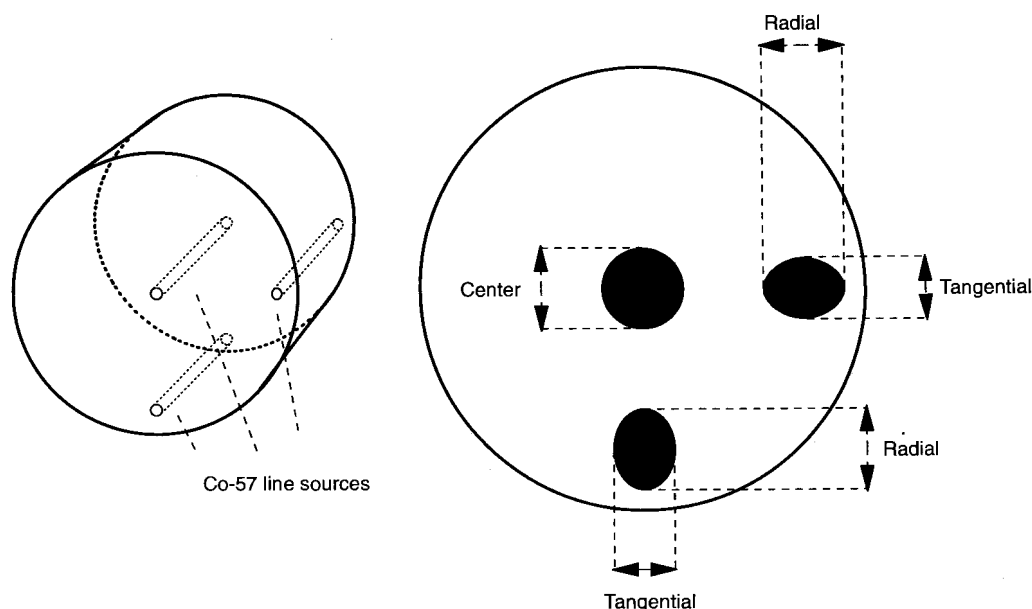


FIGURE 22-8. National Electrical Manufacturers Association (NEMA) phantom for evaluating the spatial resolution of a SPECT camera (**left**). The phantom is a 22 cm diameter plastic cylinder, filled with water and containing three Co 57 line sources. One line source lies along the central axis and the other two are parallel to the central line source, 7.5 cm away. A SPECT study is acquired with a camera radius of rotation [distance from collimator to axis of rotation (AOR)] of 15 cm. The spatial resolution is measured from reconstructed transverse images (**right**). Horizontal and vertical profiles are taken through the line sources and the full-width-at-half-maximum (FWHM) measures of these line spread functions (LSFs) are determined. The average central resolution (FWHM of the central LSF) and the average tangential and radial resolutions (determined from the FWHMs of the two peripheral sources as shown) are determined. (Adapted from Performance Measurements of Scintillation Cameras, National Electrical Manufacturers Association, 1986.)

quency cutoffs than do the ramp filters used in NEMA spatial resolution measurements. The NEMA spatial resolution measurements fail to show the advantage of SPECT systems with two or three camera heads; double- and triple-head cameras will permit the use of higher resolution collimators for clinical studies than will single-head cameras.

The spatial resolution deteriorates as the radius of the camera orbit increases. For this reason, brain SPECT produces images of much higher spatial resolution than body SPECT. It is essential that the SPECT camera heads orbit the patient as closely as possible. Also, noncircular orbits (see Design and Principles of Operation, above) provide better resolution than circular orbits.

Comparison of SPECT to Conventional Planar Scintillation Camera Imaging

In theory, SPECT should produce spatial resolution similar to that of planar scintillation camera imaging. In clinical imaging, its resolution is usually slightly worse. The camera head is usually closer to the patient in conventional planar imaging than in SPECT. Also, the short time per view of SPECT may mandate the use of a lower resolution collimator to obtain adequate numbers of counts.

In planar nuclear imaging, radioactivity in tissues in front of and behind an organ or tissue of interest causes a reduction in contrast. Furthermore, if the activity in these overlapping structures is not uniform, the pattern of this activity distribution is superimposed on the activity distribution in the organ or tissue of interest. As such, it is a source of structural noise that impedes the ability to discern the activity distribution in the organ or tissue of interest. The main advantage of SPECT over conventional planar nuclear imaging is the markedly improved contrast and reduced structural noise produced by eliminating the activity in overlapping structures. SPECT also offers the promise of partial correction for the effects of attenuation and of the scattering of photons in the patient.

Quality Control in SPECT

Although a technical quality control program is important in planar nuclear imaging, it is critical in SPECT. Equipment malfunctions or maladjustments that would not noticeably affect planar images can markedly degrade the spatial resolution of SPECT images and produce significant artifacts, some of which may mimic pathology. Upon installation, a SPECT camera should be tested by a medical physicist. Following acceptance testing, a quality control program should be established to ensure that the system's performance remains comparable to its initial performance.

X- and Y-Magnification Factors and Multienergy Spatial Registration

The X- and Y-magnification factors, often called X and Y gains, relate distances in the object being imaged, in the x and y directions, to the numbers of pixels between the corresponding points in the resultant image. The X-magnification factor is determined from a digital image of two point sources placed against the camera's collimator a known distance apart along a line parallel to the X-axis:

$$X_{\text{mag}} = \frac{\text{Actual distance between centers of point sources}}{\text{Number of pixels between centers of point sources}}$$

The Y-magnification factor is determined similarly but with the sources parallel to the y-axis. The X- and Y-magnification factors should be equal. If they are not, the projection images will be distorted in shape, as will be coronal, sagittal, and oblique images. (The transverse images, however, will not be distorted.)

The multienergy spatial registration, described in Chapter 21, is a measure of the camera's ability to maintain the same image magnification, regardless of the energies of the x- or gamma rays forming the image. It is important in SPECT when imaging radionuclides such as Ga 67 and In 111, which emit useful photons of more than one energy. It is also important because uniformity and axis of rotation corrections, to be discussed shortly, that are determined with one radionuclide will only be valid for others if the multienergy spatial registration is correct.

Alignment of Projection Images to Axis-of-Rotation (COR Calibration)

The *axis of rotation* (AOR) is an imaginary reference line about which the head or heads of a SPECT camera rotate. If a radioactive line source were placed on the AOR, each projection image would depict a vertical straight line near the center of

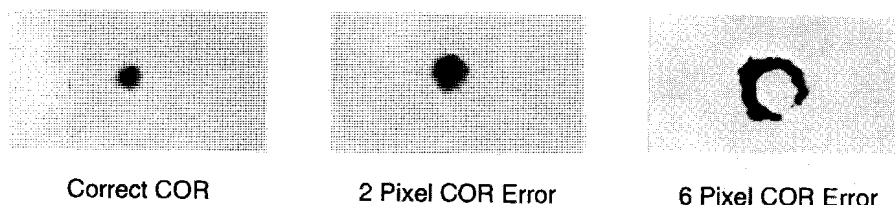


FIGURE 22-9. Center-of-rotation (COR) misalignment. Small misalignments cause blurring (**center**), whereas large misalignments cause point sources to appear as “tiny doughnut” artifacts (**right**).

the image; this projection of the AOR into the image is called the *center of rotation* (COR). Ideally, the COR is aligned with the center, in the x-direction, of each projection image. However, there may be a misalignment of the COR with the center of the projection images. This misalignment may be mechanical; for example, the camera head may not be exactly centered in the gantry. It can also be electronic. The misalignment may be the same amount in all projection images from a single camera head, or it may vary with the angle of the projection image.

If a COR misalignment is not corrected, it causes a loss of spatial resolution in the resultant transverse images. If the misalignment is large, it can cause a point source to appear as a tiny “doughnut” (Fig. 22-9). (These doughnut artifacts are not seen in clinical images; they are visible only reconstructed images of point or line sources. The doughnut artifacts caused by COR misalignment are not centered in the image and so can be distinguished from the ring artifacts caused by nonuniformities.) The COR alignment is assessed by placing a point source or line source in the camera field of view, acquiring a set of projection images, and analyzing these images using the SPECT system’s computer. If a line source is used, it is placed parallel to the AOR. The COR misalignment may be corrected by shifting each image in the x-direction by the proper number of pixels prior to filtered backprojection. When a line source is used, the COR correction can be performed separately for each slice. If the COR misalignment varies with camera head angle, instead of being constant for all projection images, it can only be corrected if the computer permits angle-by-angle corrections. Separate assessments of the COR correction must be made for different collimators and, on some systems, for different camera zoom factors and image formats (e.g., 64^2 versus 128^2). The COR correction determined using one radionuclide will only be valid for other radionuclides if the multienergy spatial registration is correct.

Uniformity

The uniformity of the camera head or heads is important; nonuniformities that are not apparent in low-count daily uniformity studies can cause significant artifacts in SPECT. The artifact caused by a nonuniformity appears in transverse images as a ring centered about the AOR (Fig. 22-10).

Multihead SPECT systems can produce partial ring artifacts when projection images are not acquired by all heads over a 360-degree arc. Clinically, ring artifacts are most apparent in high count-density studies, such as liver images. However, ring artifacts may be most harmful in studies such as Tl-201 myocardial perfusion images in which, due to poor counting statistics and large variations in count density, they may not be recognized and thus may lead to misinterpretation.

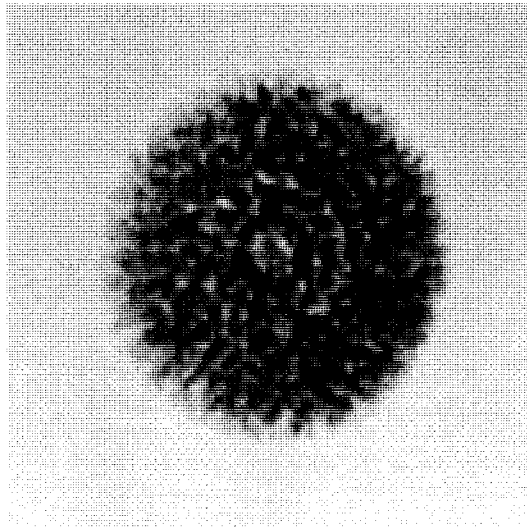


FIGURE 22-10. Image of a cylinder filled with a uniform radionuclide solution, showing a ring artifact due to a nonuniformity. Artifact is the dark ring toward the center.

As mentioned in the previous chapter, the primary intrinsic causes of nonuniformity are (a) spatial nonlinearities, which “stretch” the image in some areas, reducing the local count density, and “compress” other areas of the image, increasing the count density; and (b) local variations in the efficiency of light collection, which produce systematic position-dependent variations in the amplitudes of the energy signals and thus local variations in the fraction of counts rejected by the energy discrimination circuits. All modern SPECT cameras incorporate digital circuits that apply correction factors, obtained from internal “lookup” tables, to the X, Y, and energy signals from each interaction. However, these correction circuits cannot correct nonuniformity due to local variations in detection efficiency, such as dents or manufacturing defects in the collimators. If not too severe, nonuniformities of this type can be largely corrected. A very high count uniformity image must be acquired. The ratio of the average pixel count to the count in a specific pixel in this image serves as a correction factor for that pixel. Following the acquisition of a projection image during a SPECT study, each pixel of the projection image is multiplied by the appropriate correction factor before COR correction and filtered backprojection. For the high-count uniformity image, at least 30 million counts should be collected for 64^2 images and 120 million counts for a 128^2 format. These high-count uniformity images should be collected every 1 or 2 weeks. Separate correction images must be acquired for each camera head. For cameras from some manufacturers, separate correction images must be acquired for each radionuclide. The effectiveness of a camera’s correction circuitry and of the use of high-count flood correction images can be tested by acquiring a SPECT study of a large plastic jug filled with a well-mixed solution of Tc 99m and examining the transverse images for ring artifacts.

Camera Head Tilt

The camera head or heads must be exactly parallel to the AOR. If they are not, a loss of spatial resolution and contrast will result from out-of-slice activity being backprojected into each transverse image slice (Fig. 22-11). The loss of resolution and contrast in each transverse image slice will be less toward the center of the slice and greatest toward the edges of the image. If the axis of rotation of the camera is aligned to be level when the camera is installed and there is a flat surface on the

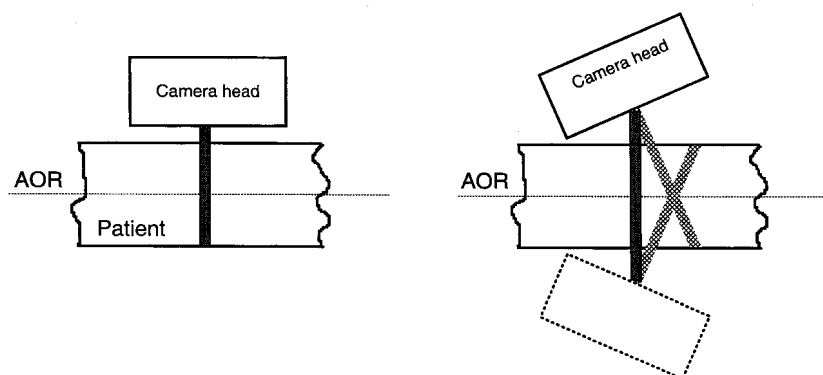


FIGURE 22-11. Head tilt. The camera head on the left is parallel to the axis of rotation (AOR), causing the counts collected in a pixel of the projection image to be backprojected into the same slice of the patient. The camera head (**right**) is tilted, causing counts from activity outside of a transverse slice (*lighter shading*) to be backprojected into the transverse slice (*darker shading*).

camera head that is parallel to the collimator face, a bubble level may be used to test for head tilt. Some SPECT cameras require the head tilt to be manually adjusted for each acquisition, whereas other systems set it automatically. The accuracy of the automatic systems should be periodically tested. A more reliable method than the bubble level is to place a point source in the camera FOV, centered in the axial (y) direction, but near the edge of the field in the transverse (x) direction. A series of projection images is then acquired. If there is head tilt, the position of the point source will vary in the y-direction from image to image. Proper head tilt is easily evaluated from cine viewing of the projection images.

SPECT Phantoms

There are commercially available phantoms (Fig. 22-12) that may be filled with a solution of Tc 99m or other radionuclide and used to evaluate system performance.

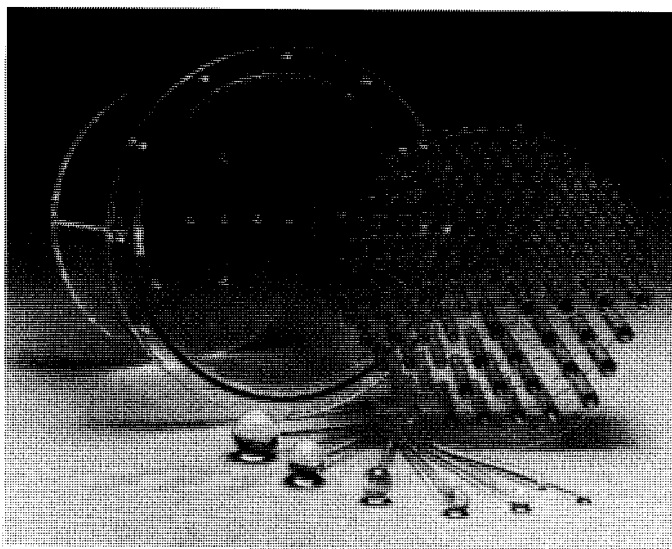


FIGURE 22-12. "Jaszczak" phantom for testing SPECT systems. (Courtesy of Data Spectrum Corp.)

TABLE 22-1. RECOMMENDED SCHEDULE FOR ROUTINE QUALITY CONTROL TESTING OF A SCINTILLATION CAMERA

Test	Frequency	Notes
Set and check energy window(s)	Daily and before imaging different radionuclide	Point source of radionuclide to be imaged
Cine review of projection images and/or review of sinogram (S) ^a	After each clinical SPECT study	Check for patient motion
Visual inspection of collimators for damage	Daily and when changing collimators	If new damage found, acquire a high-count uniformity image
Extrinsic or intrinsic low-count uniformity images of all camera heads	Daily	2 to 6 million counts, depending on effective area of camera head
High count-density extrinsic or intrinsic uniformity images of all camera heads (S)	1 or 2 weeks	30 million counts for 64 ² images and 120 million counts for 128 ²
Spatial resolution check with bar or hole pattern	1 or 2 weeks	
Center of rotation (S)	1 or 2 weeks	Point or line source, as recommended by manufacturer
Efficiency of each camera head	Quarterly	
Reconstructed cylindrical phantom uniformity (S)	Monthly	Cylindrical phantom filled with Tc-99m solution
Point source reconstructed spatial resolution (S)	Quarterly	Point source
Reconstructed SPECT phantom (S)	Quarterly	Using a phantom such as the one shown in Fig. 22-10
Pixel size check	Quarterly	Two point sources
Head tilt angle check (S)	Quarterly	Bubble level or point source
Extrinsic uniformity images of all collimators not tested above	Quarterly or semiannually	Planar source; high count density images of all collimators used for SPECT
Multienergy spatial registration	Quarterly, semiannually, or annually	Ga-67 point sources
Count rate performance	Annually	Tc-99m source

Note: The results of these tests are to be compared with baseline values, typically determined during acceptance testing. If the manufacturer recommends additional tests or more frequent testing, the manufacturer's recommendations should take precedence. For tests whose frequencies are listed as "1 or 2 weeks" or "quarterly or semiannually," it is recommended that the tests be performed initially at the higher frequency, but the frequency be reduced if the measured parameters prove stable. Tests labeled (S) need not be performed for cameras used only for planar imaging.

^aA sinogram is an image containing projection data corresponding to a single transaxial image of the patient. Each row of pixels in the sinogram is the row, corresponding to that transaxial image, of one projection image.

SPECT, single photon emission computed tomography.

These phantoms are very useful for the semiquantitative assessment of spatial resolution, image contrast, and uniformity. They are used for acceptance testing of new systems and periodic testing, typically quarterly, thereafter. Table 22-1 provides a suggested schedule for a SPECT quality-control program.

22.2 POSITRON EMISSION TOMOGRAPHY (PET)

Positron emission tomography (PET) generates images depicting the distributions of positron-emitting nuclides in patients. Figure 22-13 shows a PET scanner. In the typical scanner, several rings of detectors surround the patient. PET scanners use *annihilation coincidence detection* (ACD) instead of collimation to obtain projections of the activity distribution in the subject. The PET system's computer then reconstructs the transverse images from the projection data, as does the computer of an x-ray CT or SPECT system. Modern PET scanners are multislice devices, permitting the simultaneous acquisition of as many as 45 slices over 16 cm of axial distance. Although there are many positron-emitting radiopharmaceuticals, the clinical importance of PET today is largely due to its ability to image the radiopharmaceutical fluorine 18 fluorodeoxyglucose (FDG), a glucose analog used for differentiating malignant neoplasms from benign lesions, staging malignant neoplasms, differentiating severely hypoperfused but viable myocardium from scar, and other applications.

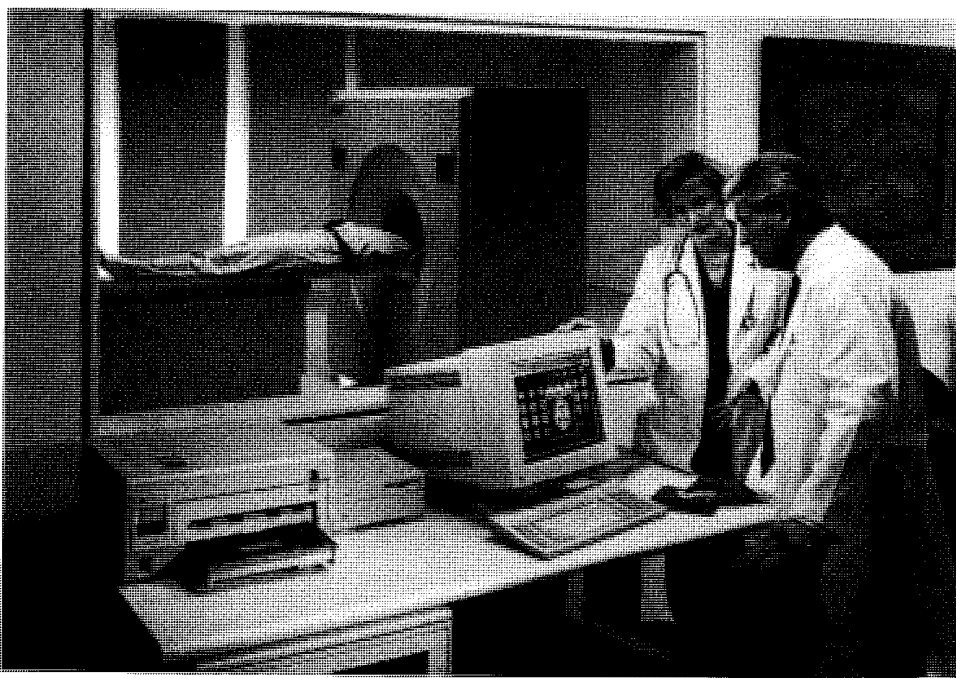


FIGURE 22-13. A commercial PET scanner. (Courtesy of Siemens Medical Systems, Nuclear Medicine Group.)

Design and Principles Of Operation

Annihilation Coincidence Detection

Positron emission is a mode of radioactive transformation and was discussed in Chapter 18. Positrons emitted in matter lose most of their kinetic energy by causing ionization and excitation. When a positron has lost most of its kinetic energy, it interacts with an electron by *annihilation* (Fig. 22-14, left). The entire mass of the electron-positron pair is converted into two 511-keV photons, which are emitted in nearly opposite directions. In solids and liquids, positrons travel only very short distances before annihilation.

If both of these annihilation photons interact with detectors, the annihilation occurred close to the line connecting the two interactions (Fig. 22-14, right). Circuitry within the scanner identifies interactions occurring at nearly the same time, a process called annihilation coincidence detection (ACD). The circuitry of the scanner then determines the line in space connecting the locations of the two interactions. Thus, ACD establishes the trajectories of detected photons, a function performed by collimation in SPECT systems. However, the ACD method is much less wasteful of photons than collimation. Additionally, ACD avoids the degradation of spatial resolution with distance from the detector that occurs when collimation is used to form projection images.

True, Random, and Scatter Coincidences

A *true coincidence* is the simultaneous interaction of emissions resulting from a single nuclear transformation. A *random coincidence* (also called an *accidental* or *chance coincidence*), which mimics a true coincidence, occurs when emissions from different nuclear transformations interact simultaneously with the detectors (Fig. 22-15). A *scatter coincidence* occurs when one or both of the photons from a single annihilation are scattered, but both are detected (Fig. 22-15). A scatter coincidence is a

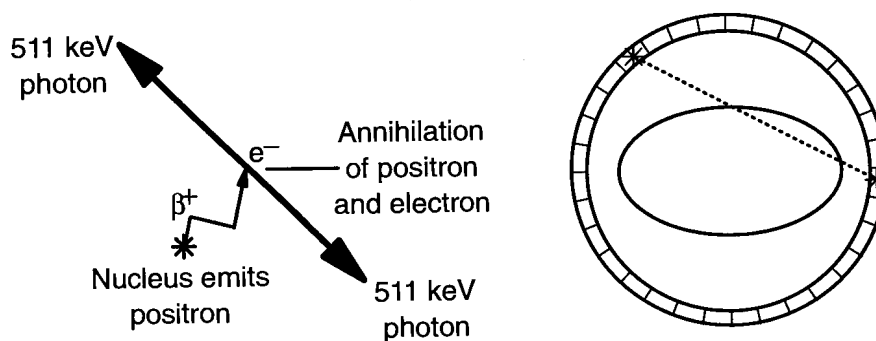


FIGURE 22-14. Annihilation coincidence detection (ACD). When a positron is emitted by a nuclear transformation, it scatters through matter losing energy. After it loses most of its energy, it annihilates with an electron, resulting in two 511-keV photons that are emitted in nearly opposite directions (**left**). When two interactions are simultaneously detected within a ring of detectors surrounding the patient (**right**), it is assumed that an annihilation occurred on the line connecting the interactions. Thus, ACD, by determining the path of the detected photons, performs the same function for the PET scanner as does the collimator of a scintillation camera.

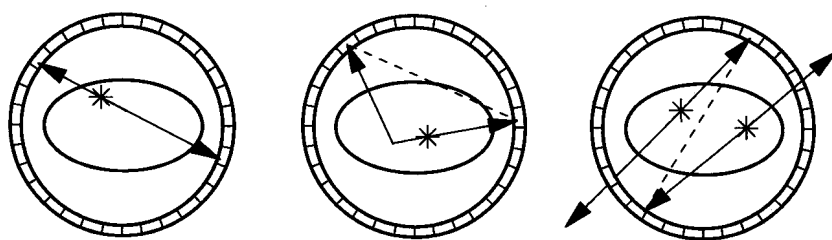


FIGURE 22-15. True coincidence (**left**), scatter coincidence (**center**), and random (accidental) coincidence (**right**). A scatter coincidence is a true coincidence, because it is caused by a single nuclear transformation, but results in a count attributed to the wrong line-of-response (*dashed line*). The random coincidence is also attributed to the wrong line of response.

true coincidence, because both interactions result from a single positron annihilation. Random coincidences and scatter coincidences result in misplaced coincidences, because they are assigned to lines of response that do not intersect the actual locations of the annihilations. They are therefore sources of noise, whose main effect is to reduce image contrast.

Design of a PET Scanner

Scintillation crystals coupled to photomultiplier tubes (PMTs) are used as detectors in PET; the low intrinsic efficiencies of gas-filled and semiconductor detectors for detecting 511-keV photons make them impractical for use in PET. The signals from the PMTs are processed using pulse mode (the signals from each interaction are processed separately from those of other interactions) to create signals identifying the position, deposited energy, and time of each interaction. The energy signal is used for energy discrimination to reduce mispositioned events due to scatter and the time signal is used for coincidence detection.

In early PET scanners, each scintillation crystal was coupled to a single PMT. In this design, the size of the individual crystal largely determined the spatial resolution of the system; reducing the size (and therefore increasing the number of crystals) improved the resolution. It became increasingly costly and impractical to pack more and more smaller PMTs into each detector ring. Modern designs couple larger crystals to more than one PMT; one such design is shown in Fig. 22-16. The relative magnitudes of the signals from the PMTs coupled to a single crystal are used to determine the position of the interaction in the crystal, as in a scintillation camera.

The scintillation material must emit light very promptly to permit true coincident interactions to be distinguished from random coincidences and to minimize dead-time count losses at high interaction rates. Also, to maximize the counting efficiency, the material must have a high linear attenuation coefficient for 511-keV photons. Most PET systems today use crystals of bismuth germanate ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$, abbreviated BGO). The light output of BGO is only 12% to 14% of that of $\text{NaI}(\text{Tl})$, but its greater density and average atomic number give it a much higher efficiency in detecting 511-keV annihilation photons. Light is emitted rather slowly from BGO (decay constant of 300 nsec), which contributes to dead-time count

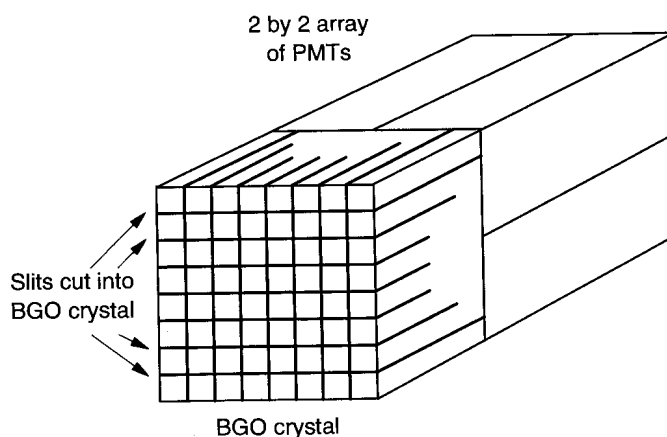


FIGURE 22-16. A technique for coupling scintillation crystals to photomultiplier tubes (PMTs). The relative heights of the pulses from the four PMTs are used to determine the position of each interaction in the crystal. The thick (3 cm) crystal is necessary to provide reasonable detection efficiency for the 511-keV annihilation photons. The slits cut in the bismuth germanate crystal form light pipes, limiting the spread of light from interactions in the front portion of the crystal, which otherwise would reduce the spatial resolution. This design permits four PMTs to serve 64 detector elements.

losses and random coincidences at high interaction rates. Several new inorganic scintillators are being investigated as possible replacements for BGO. Two of the most promising of these are lutetium oxyorthosilicate ($\text{Lu}_2\text{SiO}_4\text{O}$, abbreviated LSO) and gadolinium oxyorthosilicate ($\text{Gd}_2\text{SiO}_4\text{O}$, abbreviated GSO), both activated with cerium. Their attenuation properties are nearly as favorable as those of BGO and their much faster light emission produces better performance at high interaction rates, especially in reducing dead-time effects and in discriminating between true and random coincidences. Their higher conversion efficiencies may produce improved intrinsic spatial resolution and rejection of scatter. A disadvantage of LSO is that it is slightly radioactive; about 2.6% of naturally occurring lutetium is radioactive Lu 176, which has a half-life of 38 billion years, producing about 295 nuclear transformations per second in each cubic centimeter of LSO. NaI(Tl) is used as the scintillator in some less expensive PET systems. The properties of BGO, LSO(Ce), and GSO(Ce) are contrasted with those of NaI(Tl) in Table 22-2.

The energy signals from the detectors are sent to energy discrimination circuits, which can reject events in which the deposited energy differs significantly from 511 keV to reduce the effect of photon scattering in the patient. However, some annihilation photons that have escaped the patient without scattering interact with the detectors by Compton scattering, depositing less than 511 keV. An energy discrimination window that encompasses only the photopeak rejects these valid interactions as well as photons that have scattered in the patient. The energy window can be set to encompass only the photopeak, with maximal rejection of scatter, but also reducing the number of valid interactions detected; or the window can include part

TABLE 22-2. PROPERTIES OF SEVERAL INORGANIC SCINTILLATORS OF INTEREST IN PET

Scintillator	Decay constant (nsec)	Peak wavelength (nm)	Atomic numbers	Density (g/cm ³)	Attenuation coefficient 511 keV (cm ⁻¹)	Conversion efficiency relative to NaI
NaI(Tl)	230	415	11,53	3.67	0.343	100%
BGO	300	460	83,32,8	7.17	0.964	12–14%
GSO(Ce)	56	430	64,14,8	6.71	0.704	41%
LSO(Ce)	40	420	71,14,8	7.4	0.870	75%

Decay constants, peak wavelengths of emitted light, densities, and conversion efficiencies of BGO, GSO, and LSO from Ficke DC, Hood JT, Ter-Pogossian MM. A spheroid positron emission tomograph for brain imaging: a feasibility study. *J Nucl Med* 1996;37:1222.
 BGO, bismuth germanate; GSO, gadolinium oxyorthosilicate; LSO, lutetium oxyorthosilicate; PET, positron emission tomography.

of the Compton continuum, increasing the sensitivity, but also increasing the number of scattered photons detected.

The time signals of interactions not rejected by the energy discrimination circuits are used for coincidence detection. When a coincidence is detected, the circuitry or computer of the scanner determines a line in space connecting the two interactions (Fig. 22-14). This is called a *line of response*. The number of coincidences detected along each line of response is stored in the memory of the computer. Figure 22-17 compares the acquisition of projection data by SPECT and PET systems. Once data acquisition is complete, the computer uses the projection data to produce transverse images of the radionuclide distribution in the patient, as in x-ray CT or SPECT.

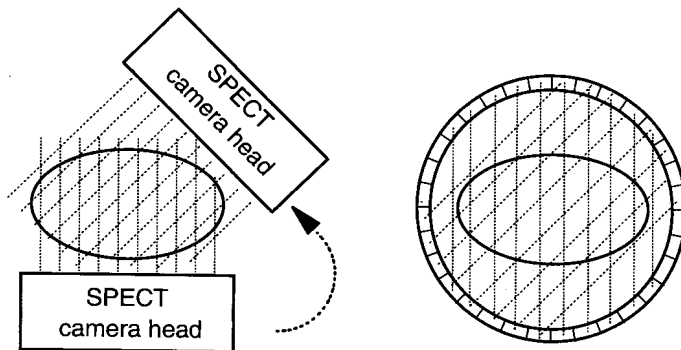


FIGURE 22-17. Acquisition of projection data by a SPECT system (**left**) and a PET system (**right**), showing how each system collects data for two projections. However, the PET system collects data for all projections simultaneously, whereas the scintillation camera head of the SPECT system must move to collect data from each projection angle.

Timing of Interactions and Detection of Coincidences

To detect coincidences, the times of individual interactions must be compared. However, interactions themselves cannot be directly detected; instead, the signals caused by interactions are detected. In scintillators, the emission of light after an interaction is characterized by a relatively rapid increase in intensity followed by a more gradual decrease. A timing circuit connected to each detector must designate one moment that follows each interaction by a constant time interval. The time signal is determined from the leading edge of the electrical signal from the PMTs connected to a detector crystal because the leading edge is steeper than the trailing edge. When the time signals from two detectors occur within a selected time interval called the *time window*, a coincidence is recorded. A typical time window for a system with BGO detectors is 12 nsec.

True Versus Random Coincidences

The rate of random coincidences between any pair of detectors is

$$R_{\text{random}} = 2\tau S_1 S_2 \quad [22-1]$$

where τ is the coincidence time window and S_1 and S_2 are the actual count rates of the detectors, often called *singles rates*. The time window is the time interval following an interaction in one detector in which an interaction in the other detector is considered to be a coincidence. The ratio of random to true coincidences increases with increasing activity in the subject and decreases as the time window is shortened. Unfortunately, there is a limit to how small the time window can be. If the time window is made too short, some true coincidences are also rejected because of imprecision in the timing of interactions. Scintillation materials that emit light promptly permit the use of shorter time windows and thus better discrimination between true and random coincidences.

Although individual random coincidences cannot be distinguished from true coincidences, methods of correction are available. The number of random coincidences along each line of response can be measured by adding a time delay to one of the two timing signals used for coincidence detection so that no true coincidences are detected. Alternatively, the number of random coincidences may be calculated from the single-event count rates of the detectors, using Equation 22-1. The number of random coincidences is then subtracted from the number of true plus random coincidences.

Scatter Coincidences

As previously mentioned, a scatter coincidence occurs when one or both of the photons from an annihilation are scattered in the patient and both are detected (Fig. 22-15). The fraction of scatter coincidences is dependent on the amount of scattering material and thus is less in head than body imaging. Because scatter coincidences are true coincidences, reducing the activity administered to the patient, reducing the time window, or using a scintillator with faster light emission does not reduce the scatter coincidence fraction. Furthermore, the corrections for random coincidences do not compensate for scatter coincidences. The energy discrimination circuits of the PET scanner reject some scatter coincidences. Unfortunately, the

relatively low conversion efficiency (Table 22-2) of BGO produces poor energy resolution, limiting the ability of energy discrimination to reject scatter. The fraction of scatter coincidences can also be reduced by two-dimensional data acquisition and axial collimation, as described in the following section. Approximate methods are available for scatter correction.

Two- and Three-Dimensional Data Acquisition

In two-dimensional (slice) data acquisition, coincidences are detected and recorded within each detector ring or small groups of adjacent detector rings. PET scanners designed for two-dimensional acquisition have thin annular collimators, typically made of tungsten, to prevent most radiation emitted by activity outside a transaxial slice from reaching the detector ring for that slice (Fig. 22-18). The fraction of scatter coincidences is greatly reduced in PET systems using two-dimensional data acquisition and axial collimation because of the geometry. Consider an annihilation occurring within a particular detector ring with the initial trajectories of the annihilation photons toward the detectors. If either of the photons scatters, it is likely that the new trajectory of the photon will cause it to miss the detector ring, thereby preventing a scatter coincidence. Furthermore, most photons from out-of-slice activity are absorbed by the axial collimators.

In two-dimensional data acquisition, coincidences within one or more pairs of adjacent detector rings may be added to improve the sensitivity (Fig. 22-19). The data from each pair of detector rings are added to that of the slice midway between the two rings. For example, if N is the number of a particular detector ring, coincidences between rings $N - 1$ and $N + 1$ are added to the projection data for ring N . For further sensitivity, coincidences between rings $N - 2$ and $N + 2$ can also be added to those of ring N . After the data from these pairs of detector rings are added, a standard two-dimensional reconstruction is performed to create a transverse image for this slice. Similarly, coincidences between two immediately adjacent detector rings, i.e., between rings N and $N + 1$, can be used to generate a transverse image

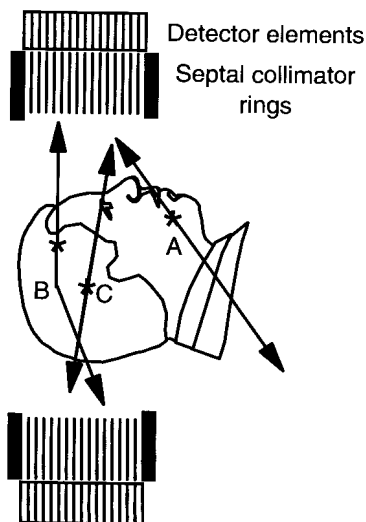


FIGURE 22-18. Side view of PET scanner illustrating two-dimensional data acquisition. The collimator rings prevent photons from activity outside the field of view (**A**) and most scattered photons (**B**) from causing counts in the detectors. However, many valid photon pairs (**C**) are also absorbed.

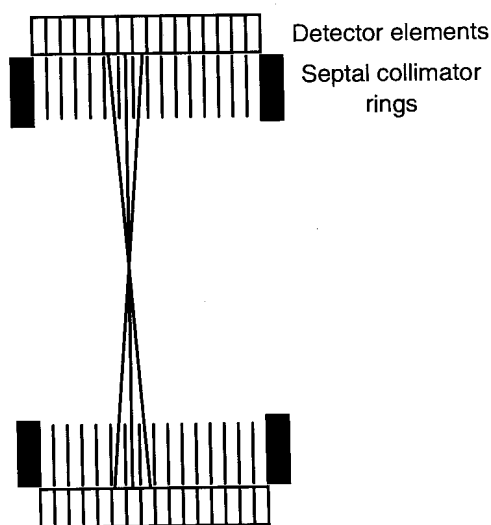


FIGURE 22-19. Side view of PET scanner performing two-dimensional data acquisition, showing lines of response for a single slice. Cross-ring coincidences have been added to those occurring within the ring of detector elements. This increases the number of coincidences detected, but causes a slight loss of axial spatial resolution that increases with distance from the axis of the scanner.

halfway between these rings. For greater sensitivity, coincidences between rings $N-1$ and $N+2$ can be added to these data. However, increasing the number of adjacent rings used in two-dimensional acquisition reduces the axial spatial resolution.

In three-dimensional (volume) data acquisition, axial collimators are not used and coincidences are detected between many or all detector rings (Fig. 22-20). Three-dimensional acquisition greatly increases the number of true coincidences detected and may permit smaller activities to be administered to patients. There are disadvantages to three-dimensional data acquisition. For the same administered activity, the greatly increased interaction rate increases the random coincidence fraction and the dead-time count losses. Thus, three-dimensional acquisition may

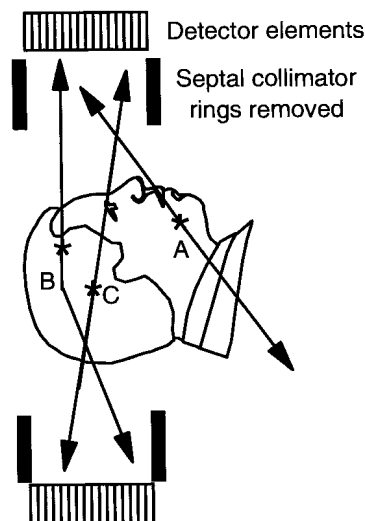


FIGURE 22-20. Side view of PET scanner illustrating three-dimensional data acquisition. Without axial collimator rings, interactions from activity outside the field of view (**A**) and scattered photons (**B**) are greatly increased, increasing the dead time, random coincidence fraction, and scatter coincidence fraction. However, the number of valid photon pairs (**C**) detected is also greatly increased.

require less activity to be administered, compared to two-dimensional image acquisition. Furthermore, the scatter coincidence fraction is much larger and the number of interactions from activity outside the FOV is greatly increased. (Activity out of the FOV causes few true coincidences, but increases the rate of random coincidences detected and dead-time count losses.) Thus, three-dimensional acquisition is likely to be most useful in low-scatter studies, such as pediatric and brain studies, and low-activity studies. Some PET systems are equipped with retractable axial collimators, permitting them to perform two- or three-dimensional acquisition.

Transverse Image Reconstruction

After the projection data are acquired, the data for each line of response is corrected for random coincidences, scatter coincidences, dead-time count losses, and attenuation (described below). Following these corrections, transverse image reconstruction is performed. For two-dimensional data acquisition, image reconstruction methods are similar to those used in SPECT. As in SPECT, either filtered backprojection or iterative reconstruction methods can be used. For three-dimensional data acquisition, special three-dimensional analytic or iterative reconstruction methods are required. These are beyond the scope of this chapter, but are discussed in the suggested readings listed at the end of the chapter.

An advantage of PET over SPECT is that, in PET, the correction for nonuniform attenuation can be applied to the projection data before reconstruction. In SPECT, the correction for nonuniform attenuation is intertwined with and complicates the reconstruction process.

Performance

Spatial Resolution

Modern whole-body PET systems achieve a spatial resolution slightly better than 5 mm FWHM of the line spread function (LSF) in the center of the detector ring when measured by the NEMA standard, *Performance Measurements of Positron Emission Tomographs*. There are three factors that primarily limit the spatial resolution of PET scanners: (a) the intrinsic spatial resolution of the detectors, (b) the distance traveled by the positrons before annihilation, and (c) the fact that the annihilation photons are not emitted in exactly opposite directions from each other. The intrinsic resolution of the detectors is the major factor determining the resolution in current scanners. In older systems in which the detectors consisted of separate crystals each attached to a single PMT, the size of the crystals determined the resolution.

The spatial resolution of a PET system is best in the center of the detector ring and decreases slightly (FWHM increases) with distance from the center. This occurs because of the considerable thickness of the detectors and because current PET systems cannot determine the depth in the crystal where an interaction occurs. Uncertainty in the depth of interaction causes uncertainty in the line of response for annihilation photons that strike the detectors obliquely. Photons emitted from the center of the detector ring can only strike the detectors head-on, but many of the photons emitted from activity away from the center strike the detectors from oblique angles (Fig. 22-21).

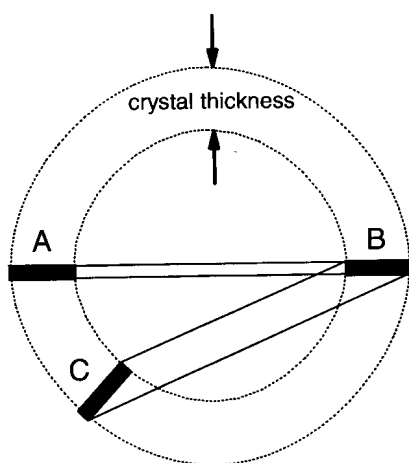


FIGURE 22-21. The cause of reduced spatial resolution with distance from the center of a PET scanner. Only three crystal segments in the detector ring are shown. For coincident interactions in detectors A and B, there is little uncertainty in the line-of-response. When coincidences occur in detectors A and C, there is greater uncertainty in the line of response. Using thinner crystals would reduce this effect, but would also reduce their intrinsic detection efficiency and thus the fraction of coincidences detected.

The distance traveled by the positron before annihilation also degrades the spatial resolution. This distance is determined by the maximal positron energy of the radionuclide and the density of the tissue. A radionuclide that emits lower energy positrons yields superior resolution. Table 22-3 lists the maximal energies of positrons emitted by radionuclides commonly used in PET. Activity in denser tissue yields higher resolution than activity in less dense tissue.

Although positrons lose nearly all of their momentum before annihilation, the positron and electron possess some residual momentum when they annihilate. Conservation of momentum predicts that the resultant photons will not be emitted in exactly opposite directions. This causes a small loss of resolution, which increases with the diameter of the detector ring.

The spatial resolution of a PET scanner, in clinical use, is likely to be somewhat worse than that measured using the NEMA protocol. The NEMA protocol specifies filtered backprojection reconstruction using a ramp filter, which provides no smoothing. Clinical projection data contains considerable statistical noise, requiring the use of a filter providing smoothing, which in turn degrades spatial resolution.

TABLE 22-3. PROPERTIES OF POSITRON-EMITTING RADIONUCLIDES COMMONLY USED IN PET

Nuclide	Half-life (minutes)	Maximal positron energy (keV)	Notes
C-11	20.5	960	
N-13	10.0	1,198	
O-15	2.0	1,732	
F-18	110	634	Most common use is ^{18}F -fluorodeoxyglucose
Rb-82	1.2	3,356	Produced by Sr-82/Rb-82 generator

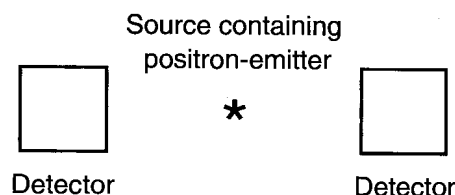


FIGURE 22-22. Point source of positron-emitting radionuclide, in air, midway between two identical detectors.

Efficiency in Annihilation Coincidence Detection

Consider a point source of a positron-emitting radionuclide in air midway between two identical detectors (Fig. 22-22) and assume that all positrons annihilate within the source. The true coincidence rate of the pair of detectors is

$$R_T = 2AG\epsilon^2 \quad [22-2]$$

where A is the activity of the source, G is the geometric efficiency of either detector, and ϵ is the intrinsic efficiency of either detector. (Geometric and intrinsic efficiency are defined in Chapter 20.) Because the rate of true coincidences detected is proportional to the square of the intrinsic efficiency, maximizing the intrinsic efficiency is very important in PET. For example, if the intrinsic efficiency of a single detector for a 511-keV photon is 0.9, 81% of the annihilation photon pairs emitted toward the detectors have coincident interactions. However, if the intrinsic efficiency of a single detector is 0.1, only 1% of pairs emitted toward the detectors has coincident interactions. As mentioned previously, most PET systems today use crystals of BGO, because of its high attenuation coefficient. These crystals are typically about 3 cm thick. As mentioned above (see Two and Three Dimensional Data Acquisition), increasing the axial acceptance angle of annihilation photons greatly increases the efficiency.

Attenuation and Attenuation Correction in PET

Attenuation in PET differs from attenuation in SPECT, because both annihilation photons must escape the patient to cause a coincident event to be registered. The probability of both photons escaping the patient without interaction is the product of the probabilities of each escaping:

$$(e^{-\mu x}) \cdot (e^{-\mu(d-x)}) = e^{-\mu d} \quad [22-3]$$

where d is the total path length through the patient, x is the distance one photon must travel to escape, and $(d-x)$ is the distance the other must travel to escape (Fig. 22-23). Thus, the probability of both escaping the patient without interaction is independent of where on the line the annihilation occurred and is the same as the probability of a single 511-keV photon passing entirely through the patient along the same path. (Equation 22-3 was derived for the case of uniform attenuation, but this principle is valid for nonuniform attenuation as well.)

Even though the attenuation coefficient for 511-keV annihilation photons in soft tissue ($\mu/\rho = 0.095 \text{ cm}^2/\text{g}$) is lower than those of photons emitted by most radionuclides used in SPECT ($\mu/\rho = 0.15 \text{ cm}^2/\text{g}$ for 140-keV gamma rays), the average path length for both to escape is much longer. For a 20-cm path in soft tissue, the chance of both annihilation photons of a pair escaping the tissue without

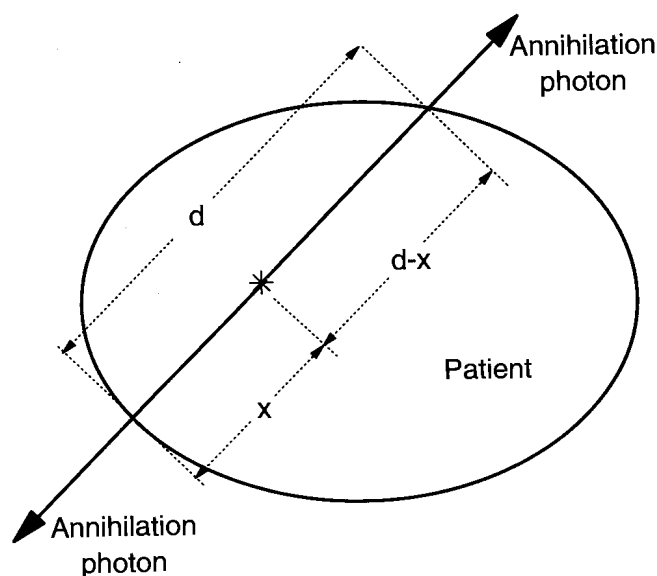


FIGURE 22-23. Attenuation in PET. The probability that both annihilation photons emitted along a particular line of response (LOR) escape interaction in the patient is independent of the position on the LOR where the annihilation occurred.

interaction is only about 15%. Thus, attenuation is more severe in PET than SPECT. The vast majority of the interactions with tissue are Compton scattering. Attenuation causes a loss of information and, because the loss is not the same for all lines of response, causes artifacts in the reconstructed transverse images. The loss of information also contributes to the statistical noise in the images.

As in SPECT, both approximate methods and methods using radioactive sources to measure the attenuation are used for attenuation correction in PET. Most of the approximate methods use a profile of the patient and assume a uniform attenuation coefficient within the profile.

Some PET systems provide one or more retractable positron-emitting sources inside the detector ring between the detectors and the patient to measure the transmission of annihilation photons from the sources through the patient. As mentioned above, the probability that a single annihilation photon from a source will pass through the patient without interaction is the same as the probability that both photons from an annihilation in the patient traveling along the same path through the patient will escape interaction. These sources are usually configured as rods and are parallel to the axis of the scanner (Fig. 22-24). The sources revolve around the patient so that attenuation is measured along all lines of response through the patient. These sources usually contain Ge 68, which has a half-life of 288 days and decays to Ga 68, which primarily decays by positron emission. Alternatively, a source containing a gamma-ray-emitting radionuclide, such as Cs 137 (662-keV gamma ray), can be used for attenuation measurements instead of a positron-emitting radionuclide source. When a gamma-ray-emitting radionuclide is used to measure the attenuation, the known position of the source and the location where a gamma ray is detected together determine a line of response for that interaction. In either case, the transmission data acquired using the external sources are used to correct the emission data for attenuation prior to transverse image reconstruction.

As in SPECT, the transmission data can be obtained before the PET radiopharmaceutical is administered to the patient, or can be obtained during emission

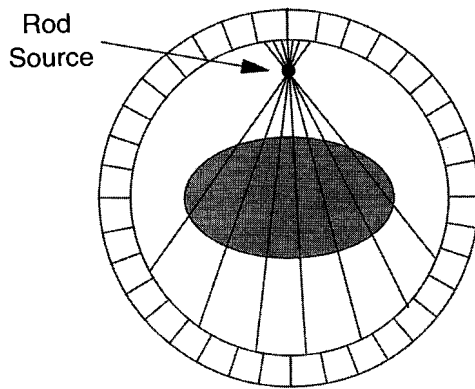


FIGURE 22-24. Rod source for attenuation correction in a dedicated PET system.

imaging. Obtaining the transmission data before the PET radiopharmaceutical is administered to the patient obviates the problem of distinguishing photons emitted by transmission source from those emitted by the radiopharmaceutical. On the other hand, if the transmission data is acquired well before the emission data, it may be difficult to adequately align the two data sets in space. For this reason, the transmission data is often acquired during the acquisition of the emission data.

There are limitations and disadvantages to attenuation correction from transmission measurements. Attenuation correction cannot compensate for the increased statistical noise caused by attenuation. Furthermore, attenuation correction increases the imaging time, although only slightly if performed during emission image acquisition; slightly increases the radiation dose to the patient; and increases the statistical noise in the images. Also, spatial misregistration between the attenuation map and the emission data, most often caused by patient motion, can cause significant artifacts. Attenuation correction may reduce the sensitivity of some clinical studies. Attenuation correction may not be necessary for all PET procedures and, for some procedures in which it is performed, the interpreting physician should review the uncorrected reconstructed images as well.

Alternatives to Dedicated PET Systems—SPECT with High-Energy Collimators and Multihead SPECT Cameras with Coincidence Detection Capability

Some nuclear medicine departments may wish to image positron-emitting radiopharmaceuticals, but lack sufficient demand to justify a dedicated PET scanner. Positron-emitting radiopharmaceuticals also can be imaged using standard scintillation cameras with ultra-high-energy collimators or multihead SPECT cameras with coincidence detection capability.

Some scintillation cameras are capable of performing planar imaging and SPECT of positron emitters using collimators designed for 511-keV photons. These collimators have very thick septa and have very low collimator efficiencies. Despite the thick septa, the images produced suffer from considerable septal penetration. Also, the thick septa produce collimator pattern artifacts in the images. However, attenuation by the patient is much less significant than in coincidence imaging. The gantries of such cameras must be able to support very heavy camera heads, because of the weight of

511-keV collimators and because the camera heads must contain adequate shielding for 511-keV photons from outside the field of view.

Double- and triple-head SPECT cameras with coincidence detection capability are commercially available. These cameras have planar NaI(Tl) crystals coupled to PMTs. With standard nuclear medicine collimators mounted, they function as standard scintillation cameras and can perform planar nuclear medicine imaging and SPECT. With the collimators removed or replaced by axial collimators, they can perform coincidence imaging of positron-emitting radiopharmaceuticals.

When used for coincidence imaging, these cameras provide spatial resolution similar to that of a dedicated PET system. However, in comparison to a dedicated PET system, they detect a much lower number of true coincidences. The smaller attenuation coefficient of sodium iodide and the limited crystal thickness limits the intrinsic efficiency of each detector. For example, the photopeak efficiency of a 0.95 cm (3/8 inch) thick NaI(Tl) crystal for annihilation photons is about 12% (see Chapter 21, Fig. 21-9). The fraction of annihilation photon pairs incident on a pair of detectors that yields true coincidences is equal to the square of the photopeak efficiency (assuming that only photopeak-photopeak coincidences are registered). Thus, for 0.95-cm thick crystals, only about 1.5% of annihilation photon pairs reaching the detectors are registered as true coincidences. For this reason, many such cameras have thicker NaI crystals than standard scintillation cameras, with thicknesses ranging from 1.27 to 2.5 cm (1/2 to 1 inch). Even modest increases in crystal thickness significantly improve the number of true coincidences detected. However, even with thicker crystals, the coincidence detection sensitivities of multihead scintillation cameras with coincidence circuitry are much lower than those of dedicated PET systems. The lower coincidence detection sensitivity causes more statistical noise in the images than in images from a dedicated PET system. Furthermore, the low coincidence detection sensitivity forces the use of measures, such as wide spacing of axial septa in two-dimensional acquisition or three-dimensional acquisition and wide energy discrimination windows, to obtain adequate numbers of true coincidences. These measures increase the scatter coincidence fraction, producing images of lower contrast than a dedicated PET system. Both the greater noise due to fewer coincidences and the reduced contrast due to the higher scatter coincidence fraction limit the ability to detect small lesions in the patient, in comparison to a dedicated PET system. When these cameras are used with collimators as standard scintillation cameras, their thicker crystals produce slightly worse intrinsic spatial resolution for lower energy photons, such as those emitted by Tc 99m and Tl 201, but have little effect on the system spatial resolution.

Some manufacturers provide two-dimensional data acquisition, whereas others provide three-dimensional acquisition (described earlier in this chapter). The cameras using two-dimensional data acquisition are provided with axial collimators to reduce the singles rates from out-of-slice activity and to reduce the fraction of scatter coincidences.

Because these cameras are either uncollimated or used with axial collimators when performing coincidence imaging, they are subject to very high interaction rates and must have very fast electronics. Because such a system has only two or three detectors (scintillation camera heads) instead of the hundreds of detectors in a dedicated PET system, its performance suffers more from dead-time count losses.

Although many interactions from lower energy photons are rejected by the energy discrimination circuits of the camera, they still cause dead-time count losses and can interfere with the determination of the energies and positions of interactions.

Some cameras are equipped with filters, between the crystal and the axial collimator in two-dimensional mode and in place of the collimator in three-dimensional mode, to preferentially absorb low-energy photons. These filters typically have several layers, such as lead, tin, and copper, with the atomic numbers of the layers decreasing toward the crystal to attenuate characteristic x-rays from the higher atomic number layers.

Clinical experience is accumulating regarding the utility of imaging F-18 FDG using scintillation cameras with either 511-keV collimators or coincidence detection capability. It appears that SPECT with high-energy collimators is acceptable for the determination of myocardial viability, although less accurate than a dedicated PET system. However, attenuation artifacts pose a problem in determining myocardial viability by coincidence imaging with a scintillation camera and accurate attenuation correction may be necessary for this application.

F-18 FDG imaging with high-energy collimators does not appear to be acceptable for brain imaging or for the evaluation and staging of neoplasms, because of its limited efficiency and spatial resolution. On the other hand, coincidence imaging with a scintillation camera may be useful for the detection and staging of neoplasms, but is less sensitive than a dedicated PET system, particularly for detecting smaller tumors. The images from coincidence imaging with a scintillation camera suffer from more statistical noise and have less contrast than those from a dedicated PET system. Coincidence imaging with scintillation cameras for oncologic applications is likely to be more effective in imaging the head, neck, and thorax, where the effects of attenuation and scattering are less severe than in the abdomen and pelvis.

Economical Dedicated PET Systems

Several vendors have developed dedicated PET systems that are less expensive than PET systems with full rings of BGO detectors, but that provide much better coincidence detection sensitivity than double-head scintillation camera coincidence systems. One such design consists of six stationary uncollimated scintillation cameras, each with a curved 2.5 cm thick NaI(Tl) crystal, surrounding the patient (Fig. 22-25, left).

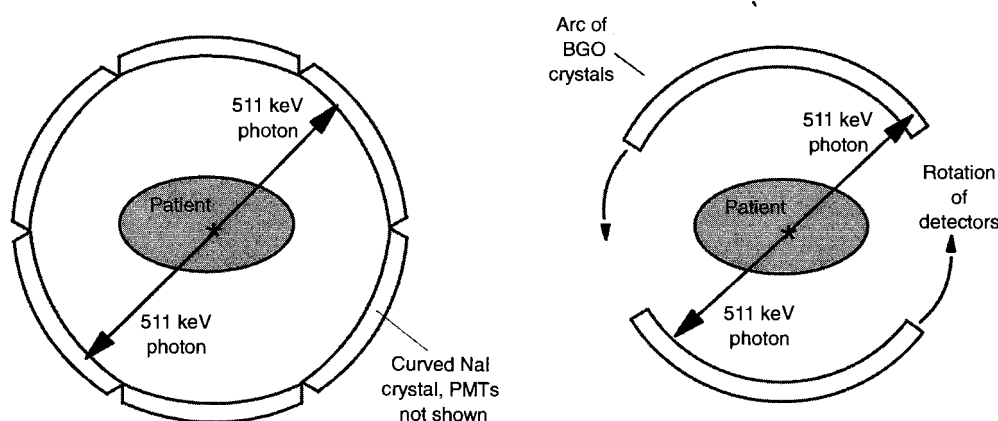


FIGURE 22-25. Less expensive dedicated PET systems. **Left:** This system consists of six scintillation cameras, each with a curved 2.5 cm thick NaI(Tl) crystal, facing the patient. The camera heads are stationary. **Right:** This system consists of two opposed arcs of BGO detectors, of the type shown in Fig. 22-14, facing the patient. To obtain projection data from all lines of response, the detectors revolve around the patient.

A system provided by another vendor consists of two opposed arcs of BGO detectors (Fig. 22-25, right). To obtain projection data over a full 180 degrees around the patient, the detectors revolve about the patient. Both systems use coincidence detection in lieu of collimation to obtain projection data.

Comparison of SPECT and PET

In single photon emission imaging, the spatial resolution and the detection efficiency are primarily determined by the collimator. Both are ultimately limited by the compromise between collimator efficiency and collimator spatial resolution that is a consequence of collimated image formation. It is the use of annihilation coincidence detection instead of collimation that makes the PET scanner much more efficient than the scintillation camera and also yields its superior spatial resolution.

In systems that use collimation to form images, the spatial resolution rapidly deteriorates with distance from the face of the imaging device. This causes the spatial resolution to deteriorate from the edge to the center in transverse SPECT images. In contrast, the spatial resolution in a transverse PET image is best in the center. Table 22-4 compares SPECT and PET systems.

TABLE 22-4. COMPARISON OF SPECT AND PET

	SPECT	PET
Principle of projection data collection	Collimation	Annihilation coincidence detection
Transverse image reconstruction	Filtered backprojection or iterative methods	Filtered backprojection or iterative methods
Radionuclides	Any emitting x-rays, gamma rays, or annihilation photons; optimal performance for photon energies of 100 to 200 keV.	Positron emitters only
Spatial resolution	Depends on collimator and camera orbit Within a transaxial image, the resolution in the radial direction is relatively uniform, but the tangential resolution is degraded toward the center Typically about 10-mm FWHM at center for a 30-cm diameter orbit and Tc-99m (NEMA protocol) Larger camera orbits produce worse resolution	Relatively constant across transaxial image, best at center Typically 4.5 to 5 mm FWHM at center (NEMA protocol)
Attenuation	Attenuation less severe Attenuation correction sources available, but utility not yet established	Attenuation more severe However, accurate attenuation correction is possible, with a transmission source
Cost	Typically \$500,000 for a double-head, variable-angle camera	\$1,000,000 to \$2,000,000

FWHM, full width at half maximum.

Factors Affecting the Availability Of PET

The main factors limiting the availability of PET are the relatively high cost of a dedicated PET scanner and, in many areas of the country, the lack of local sources of F 18 FDG. However, cyclotron-equipped commercial radiopharmacies are now located in many major metropolitan areas. Multihead SPECT cameras with coincidence circuitry and SPECT cameras with high-energy collimators provide less expensive, although less accurate, alternatives for imaging FDG. PET imaging using FDG has become routine at many major medical centers.

Combined PET/X-Ray CT and SPECT/Coincidence Imaging/X-Ray CT Systems

PET imaging, particularly in oncologic applications, often reveals suspicious lesions, but provides little information regarding their exact locations in the organs of the patients. In such situations, co-registration of the PET images with images from another modality, such as x-ray CT or magnetic resonance imaging (MRI), that provides good depiction of anatomy, can be very useful. However, this usually requires moving the patient from the PET system to the CT or MRI system and, once the patient has been moved, it is very difficult to accurately align the PET data with the CT or MRI data. To solve this problem, systems have been developed that incorporate a PET system and a conventional x-ray CT system in a single gantry. The patient lies on a bed that passes through the bores of both systems. A series of CT images is acquired over the same section of the patient as the PET scan, either before or after the PET image acquisition. Because the bed moves in the axial direction, but the patient lies still on the bed, each CT image slice corresponds to a transverse PET image, permitting accurate co-registration. Furthermore, the x-ray CT information can be used to provide attenuation correction of the PET information. In the co-registered images, the PET information is usually superimposed in color on gray-scale CT images.

One manufacturer provides a gantry containing two scintillation camera heads and a simple x-ray CT machine. The scintillation camera heads can be used for SPECT imaging of conventional radiopharmaceuticals and, when placed in the 180-degree orientation, for coincidence imaging of positron-emitting radiopharmaceuticals. The x-ray CT device provides CT image data for image co-registration and for attenuation correction.

SUGGESTED READING

- Bax JJ, Wijns W. Editorial: Fluorodeoxyglucose imaging to assess myocardial viability: PET, SPECT, or gamma camera coincidence imaging. *J Nucl Med* 1999;40:1893–1895.
- Cherry SR, Phelps ME. Positron emission tomography: methods and instrumentation. In: Sandler MP, et al., eds. *Diagnostic nuclear medicine*, 3rd ed., vol. 1. Baltimore: Williams & Wilkins, 1996:139–159.
- Gelfand MJ, Thomas SR. *Effective use of computers in nuclear medicine*. New York: McGraw-Hill, 1988.
- Groch MW, Erwin WD. SPECT in the Year 2000: Basic principles. *J Nuclear Med Tech* 2000;28:233–244.
- Groch MW, Erwin WD, Bieszk JA. Single photon emission computed tomography. In: Treves ST, ed. *Pediatric nuclear medicine*, 2nd ed. New York: Springer-Verlag, 1995:33–87.

- Madsen MT. The AAPM/RSNA physics tutorial for residents. Introduction to emission CT. *RadioGraphics* 1995;15:975–991.
- Miller TR. The AAPM/RSNA physics tutorial for residents. Clinical aspects of emission tomography. *RadioGraphics* 1996;16:661–668.
- NEMA. Recommendations for implementing SPECT instrumentation quality control. *J Nucl Med* 2000;41:383–389.
- Votaw JR. The AAPM/RSNA physics tutorial for residents. Physics of PET. *RadioGraphics* 1995;15:1179–1190.
- Yester MV, Graham LS, eds. *Advances in nuclear medicine: The medical physicist's perspective*. Proceedings of the 1998 Nuclear Medicine Mini Summer School, American Association of Physicists in Medicine, June 21–23, 1998, Madison, WI.

SECTION
IV

**RADIATION PROTECTION,
DOSIMETRY, AND BIOLOGY**

RADIATION PROTECTION

It is incumbent upon all individuals who use radiation in medicine to strive for an optimal compromise between its clinical utility and the radiation dose to patients, staff, and the public. To a large degree, the success of radiation protection programs depends on the education of staff about radiation safety principles and the risks associated with radiation exposure and contamination. This chapter discusses the application of radiation protection principles (also known as health physics) in diagnostic radiology and nuclear medicine.

23.1 SOURCES OF EXPOSURE TO IONIZING RADIATION

According to the National Council on Radiation Protection and Measurement (NCRP) report no. 93, the annual average per capita total effective dose equivalent (exclusive of smoking) from exposure to ionizing radiation in the United States is approximately 3.6 millisievert (mSv) (360 mrem). Approximately 80% of this exposure, ~3 mSv (300 mrem), is from naturally occurring sources, whereas 20%, ~600 μ Sv (60 mrem), is from technologic enhancements of naturally occurring sources and artificial radiation sources, the majority of which are diagnostic x-ray procedures. These averages apply to the population of the United States as a whole. Doses to individuals from these sources vary with a variety of factors discussed below.

Naturally Occurring Radiation Sources

Naturally occurring sources of radiation include (a) cosmic rays, (b) cosmogenic radionuclides, and (c) primordial radionuclides and their radioactive decay products. Cosmic radiation includes both the primary extraterrestrial radiation that strikes the earth's atmosphere and the secondary radiations produced by the interaction of primary cosmic rays with the atmosphere. Primary cosmic rays predominantly consist of extremely penetrating high-energy (mean energy ~10 BeV) particulate radiations. Almost all cosmic radiation (approximately 80% of which is high-energy protons) collides with our atmosphere, producing showers of secondary particulate radiations (e.g., electrons and muons) and electromagnetic radiation. The average per capita equivalent dose from cosmic radiation is ~270 μ Sv (27 mrem) per year or ~9% of natural background radiation. However, it is important to note that the range of individual exposure is considerable. The majority of the population of the United States

is exposed to cosmic radiation near sea level where the dose equivalent rate is ~ 240 $\mu\text{Sv}/\text{year}$ (24 mrem/year). However, smaller populations receive more than five times this amount [e.g., Leadville, Colorado, at 3,200 m, ~ 1.25 mSv/year (125 mrem/year)]. Exposures increase with altitude, approximately doubling every 1,500 m, as there is less atmosphere to attenuate the cosmic radiation. Cosmic radiation is also greater at the earth's poles than at the equator, as charged particles encountered by the earth's magnetic field are forced to travel along the field lines to either the North or South Pole. Structures provide some protection from cosmic radiation; the indoor effective dose equivalent rate is $\sim 20\%$ lower than outdoors.

Air travel can substantially add to an individual's cosmic ray exposure. For example, airline crews and frequent fliers receive an additional annual equivalent dose on the order of ~ 1 mSv (100 mrem); some receive an equivalent dose several times higher. A 5-hour transcontinental flight in a commercial jet aircraft will result in an equivalent dose of ~ 25 μSv (2.5 mrem). Supersonic high altitude aircraft, such as the British Concord, have on-board radiation monitors to alert the crews to descend to lower altitudes if excessive cosmic radiation dose rates are detected. Spacecraft also experience higher exposures to cosmic rays; the Apollo astronauts received an average equivalent dose of 2.75 mSv (275 mrem) during a lunar mission.

A fraction of the secondary cosmic ray particles collide with stable atmospheric nuclei producing "cosmogenic" radionuclides (e.g., $^{14}_7\text{N}$ [n,p] $^{14}_6\text{C}$). Although many types of cosmogenic radionuclides are produced, they contribute very little (~ 4 $\mu\text{Sv}/\text{year}$ [~ 0.4 mrem/year] or $< 1\%$) to natural background radiation. The majority of the effective dose equivalent caused by cosmogenic radionuclides is from carbon 14.

The terrestrial radioactive materials that have been present on earth since its formation are called *primordial radionuclides*. Primordial radionuclides with physical half-lives comparable to the age of the earth (~ 4.5 billion years) and their radioactive decay products are the largest sources of terrestrial radiation exposure. The population radiation dose from primordial radionuclides is the result of external radiation exposure, inhalation, and incorporation of radionuclides in the body. Primordial radionuclides with half-lives less than 10^8 years have decayed to undetectable levels since their formation, whereas those with half-lives greater than 10^{10} years do not significantly contribute to background radiation levels because of their long physical half-lives (i.e., slow rates of decay). Most radionuclides with atomic numbers greater than lead decay to stable isotopes of lead through series of radionuclide decays called *decay chains*. The radionuclides in these decay chains have half-lives ranging from seconds to many thousands of years. Other primordial radionuclides, such as potassium 40 (K-40, $T_{1/2} = 1.29 \times 10^9$ years) decay directly to stable nuclides. The decay chains of uranium 238 (U-238), $T_{1/2} = 4.51 \times 10^9$ years (uranium series), and thorium 232 (Th-232), $T_{1/2} = 1.41 \times 10^{10}$ years (thorium series), produce several dozen radionuclides that together with K-40 are responsible for the majority of the external terrestrial average equivalent dose rate of 280 $\mu\text{Sv}/\text{year}$ (28 mrem/year) or $\sim 9\%$ of natural background. Individuals may receive higher or lower exposures than the average, depending on the local concentration of terrestrial radionuclides. The range in the United States is ~ 0.15 to 25 mSv/year (15 to 2,500 mrem/year), and there are a few regions of the world where terrestrial radiation sources are highly concentrated. For example, in Brazil, monazite sand deposits, containing high concentrations of radionuclides from the Th-232 decay series, are found along certain beaches. The external radiation levels on these black sands range up to 50 $\mu\text{Gy}/\text{hr}$

(5 mrad/hr), which is approximately 400 times the average total background in the U.S., ~ 0.011 mSv/hr (1.1 mrem/hr), excluding radon.

The short-lived decay products of radon 222 (Rn-222) are the most significant source of exposure from the inhalation of naturally occurring radionuclides. Radon-222, a noble gas, is produced in the U-238 decay chain by the decay of radium-226 (Ra-226). Rn-222 decays by alpha emission, with a half-life of 3.8 days, to polonium 218 (Po-218), followed by several other alpha and beta decays, eventually leading to stable lead-206 (Pb-206). Once the short-lived daughters of radon are inhaled, the majority of the dose is deposited in the tracheobronchial region of the lung. Radon concentrations in the environment vary widely. There are both seasonal and diurnal variations in radon concentrations. Radon gas emanates primarily from the soil in proportion to the quantity of natural uranium deposits; its dispersion is restricted by structures producing significantly higher levels than found outdoors in the same area. Weatherproofing of homes and offices and other energy conservation techniques typically decrease outside air ventilation, resulting in higher indoor radon concentrations.

The exposure from Rn-222 in the United States results in an average equivalent dose to the bronchial epithelium of 24 mSv/year (2.4 rem/year). Assuming a tissue weighting factor of 0.08 for the bronchial epithelium yields an effective dose equivalent rate of ~ 2 mSv/year (200 mrem/year) or $\sim 68\%$ of natural background. The average indoor air concentration of Rn-222 in the United States is ~ 55 Bq/m³ (1.5 pCi/L); however, levels can exceed 2.75 kBq/m³ (75 pCi/L) in poorly ventilated structures with high concentrations of U-238 in the soil. Outdoor air concentrations are ~ 10 times lower, 5.5 Bq/m³ (0.15 pCi/L). The U.S. Environmental Protection Agency (EPA) and the NCRP recommend taking action to reduce radon levels in homes exceeding 147 and 294 Bq/m³ (4 and 8 pCi/L), respectively, whereas other countries have different limits [e.g., the United Kingdom and Canada have higher action levels, 733 Bq/m³ (20 pCi/L)]. Although Rn-222 accounts for about two-thirds of natural radiation exposure, it can be easily measured and exposures can be reduced when necessary.

The second largest source of natural background radiation is from ingestion of food and water containing primordial radionuclides (and their decay products), of which K-40 is the most significant. K-40 is a naturally occurring isotope of potassium ($\sim 0.01\%$). Skeletal muscle has the highest concentration of potassium in the body. K-40 and to a lesser extent other dietary sources of primordial radionuclides produce an average effective dose equivalent rate of ~ 400 μ Sv/year (40 mrem/year), or $\sim 13\%$ of natural background. Table 23-1 lists these sources and their contributions to natural background radiation.

Technology-Based Radiation Exposure

Modern technology has increased population radiation exposure. Technology-based population exposure to radiation sources can be categorized as (a) enhanced natural sources, in which technologic enhancements to naturally occurring radiation sources have resulted in higher population exposures than would have occurred if left in their natural state; and (b) technologically produced (i.e., artificial or man-made) radiation sources. With the exception of smoking, these sources represent less than one-fourth, ~ 700 μ Sv/year, (70 mrem/year) of the dose from natural background.

TABLE 23-1. ESTIMATED AVERAGE TOTAL EFFECTIVE DOSE EQUIVALENT RATE IN THE UNITED STATES FROM VARIOUS SOURCES OF NATURAL BACKGROUND RADIATION

	Source	Type of exposure; tissue exposed	Annual effective dose equivalent		% Total
			mSv	mrem	
Primordial radionuclides	Inhaled radionuclides (primarily Rn-222 and daughter products)	Internal; bronchial epithelium	2	200	68
	Ingested radionuclides (primarily K-40)	Internal; whole body	0.4	40	13
	Exposure from external terrestrial radionuclides	External; whole body	0.28	28	9
	Cosmic rays	External; whole body	0.27	27	9
	Cosmogenic radionuclides	Internal; whole body	<0.01	<1	<1
Total			~3	~300	100

Adapted from National Council on Radiation Protection and Measurements. *Exposure of the population in the United States and Canada from natural background radiation*. NCRP report no. 94. Bethesda, MD: National Council on Radiation Protection and Measurements, 1988.

Enhanced Natural Sources

This category includes a variety of sources, most of which are consumer products. The largest contribution in this category is from tobacco products, which are estimated to produce an annual equivalent dose to the bronchial epithelium of ~160 mSv (16 rem) for smokers. Utilizing a tissue weighting factor of 0.08 yields an annual effective dose equivalent rate of ~13 mSv (1.3 rem) that, if averaged over the U.S. population (smokers and nonsmokers), would produce an annual average effective dose equivalent rate of ~2.8 mSv (280 mrem). The appropriateness of this conversion from equivalent dose to effective dose equivalent for this source is open to question because only a small percentage of the bronchial epithelium is actually exposed. For this reason and other uncertainties in the assumptions needed to estimate population doses, the NCRP describes but does not include this source in their estimate of the average annual effective dose equivalent.

Radon gas dissolved in domestic water supplies can be released into the air within a home during water usage. This source is primarily associated with well and other ground water sources. The NCRP estimates this source to contribute from 10 to 60 μ Sv (1 to 6 mrem) to the annual average effective dose equivalent rate in the United States.

Many building materials contain radioisotopes of uranium, thorium, and potassium. These primordial radionuclides and their decay products are found in higher concentration in such materials as brick, concrete, and granite, and thus structures built from these materials will cause higher exposures than structures constructed primarily from wood. The annual average per capita effective dose equivalent rate is estimated to be ~30 μ Sv (3 mrem) from this source.

There are numerous other less important sources of enhanced natural radiation exposure, such as mining and agricultural activities (primarily from fertilizers containing members of the uranium and thorium decay series and K-40);

combustible fuels, including coal and natural gas (radon); and consumer products, including smoke alarms (americium-241); gas lantern mantles (thorium); dental prostheses, certain ceramics, and optical lenses (uranium). These sources contribute <1% of the average annual effective dose equivalent from enhanced natural sources.

Artificial Sources

The greatest exposure to the U.S. population from artificial radiation sources is from medical diagnosis and therapy. The majority of the exposure is from medical x-rays (primarily from diagnostic radiology), with a smaller contribution from nuclear medicine. Doses from individual medical examinations are discussed in detail in Chapter 24. The medical use of radiation produces an annual average effective dose equivalent of $\sim 540 \mu\text{Sv}$ (54 mrem), which represents more than 95% of the total from artificial radiation sources and $\sim 88\%$ of all technology based radiation sources (excluding smoking). The data on medical radiation exposure were taken from the most recent information published by the NCRP (report no. 100, 1989). There have been significant changes in medical imaging technology and its uses since these data were collected. The dose per image from radiographic procedures has substantially decreased. However, the doses from fluoroscopy and computed tomography (CT) have increased since that time.

Fallout, from the atmospheric testing of nuclear weapons (450 detonations between 1945 and 1980), consists of large quantities of carbon-14 (70% of the total fallout) and a variety of other radionuclides including tritium (H-3), manganese-54 (Mn-54), cesium-136 and 137 (Cs-136, Cs-137), barium-140 (Ba-140), cerium-144 (Ce-144), plutonium, and transplutonium elements. A large fraction of these radionuclides have since decayed and/or have become progressively less available for biologic uptake. Today, the dose from fallout is minimal and primarily due to carbon-14, which, together with other fallout radionuclides, results in an annual average effective dose equivalent of $<10 \mu\text{Sv}$ (1 mrem).

The total contribution to the annual effective dose equivalent from activities related to nuclear power production (known as the *nuclear fuel cycle*) is minimal, $<0.5 \mu\text{Sv}$ (0.05 mrem). Population radiation exposure from nuclear power production occurs during all phases of the fuel cycle, including mining and manufacturing of uranium fuel, reactor operations, and radioactive waste disposal. Although many radionuclides are produced during the nuclear fuel cycle, C-14 is the most significant contributor to population effective dose equivalent.

Table 23-2 shows the contributions of these sources to the average annual effective dose equivalent. It is interesting to note that although nuclear power, fallout from atomic weapons, and a variety of consumer products such as video display terminals receive considerable attention in the popular media, they in fact cause only a small fraction of the average population exposure to radiation.

Occupational Exposures

On average, occupational exposures associated with uranium mining are among the highest, $\sim 12 \text{ mSv/year}$ (1.2 rem/year). Other occupational exposures occur in nuclear power operations, medical diagnosis and therapy, aviation and such miscel-

TABLE 23-2. ESTIMATED AVERAGE ANNUAL TOTAL EFFECTIVE DOSE EQUIVALENT IN THE UNITED STATES FROM TECHNOLOGICALLY ENHANCED AND PRODUCED RADIATION SOURCES

Source	Type of exposure location tissue exposed	Annual effective dose equivalent		% Total (Exclusive of smoking)
		mSv	mrem	
Enhanced natural sources				
Smoking (Po 210)	Internal; bronchial epithelium	2.80 (13) ^a	280 (1,300)	NA
Domestic water supply (radon)	Internal; bronchial epithelium	0.03	3	5
Building materials	Mixed external and internal; multiple sites	0.03	3	5
Mining and agricultural products	Mixed external and internal; multiple sites	<0.01	<1	<1
Combustible fuels	Mixed external and internal; multiple sites	<0.01	<1	<1
Other consumer products	Mixed external and internal; multiple sites	<0.01	<1	<1
Subtotal		~0.07	~7	12%
Technologically produced (artificial or man-made)				
Medical x-rays	External, multiple sites	0.39	39	65
Nuclear medicine	Internal, multiple sites	0.14	14	24
Fallout	Mixed external and internal; multiple sites	0.01	<1	<1
Misc environmental sources	Mixed external and internal; multiple sites	<<0.01	0.06	<<1
Nuclear power	Mixed external and internal; multiple sites	<<0.01	<0.05	<<1
Subtotal		~0.54	~54	~88%
Total (Exclusive of smoking)^b		~0.60	~60	~100%

Adapted from National Council on Radiation Protection and Measurements. *Radiation exposure of the U.S. population from consumer products and miscellaneous sources*. NCRP report no. 95. Bethesda, MD: National Council on Radiation Protection and Measurements, 1988.

^aThe figure in parenthesis (13) is for smokers only; the other figure [2.80] represents the dose to smokers averaged over the total U.S. population.

^bDue to uncertainties involved in converting dose to effective dose for this exposure, the National Council on Radiation Protection (NCRP) does not currently recommend including this source in their estimates of population average annual effective dose equivalent.

laneous activities as research, nonuranium mining, and application of phosphate fertilizers (Table 23-3).

Physicians and x-ray technologists in diagnostic radiology receive an average annual effective dose equivalent of approximately 1 mSv (100 mrem). This average is somewhat lower than might be expected because it includes personnel who receive very small occupational exposures (e.g., radiology supervisors). Annual personnel equivalent doses exceeding 15 mSv (1,500 mrem) are typical of special procedures utilizing fluoroscopy and cineradiography (e.g., cardiac catheterization). In reality, these are only partial body exposures (i.e., head and extremities) because the use of lead aprons greatly reduces the exposure to most of the body. It is interesting to note, however, that the

TABLE 23-3. AVERAGE ANNUAL OCCUPATIONAL EFFECTIVE DOSE EQUIVALENT IN THE UNITED STATES

Occupational category	Average annual total effective dose equivalent	
	mSv	mrem
Uranium miners ^a	12.0	1,200
Nuclear power operations ^b	6.0	600
Airline crews	1.7	170
Diagnostic radiology and nuclear medicine techs	1.0	100
Radiologists	0.7	70

Adapted for measurably exposed personnel from National Council on Radiation Protection and Measurements. *Exposure of the U.S. population from occupational radiation*. NCRP report no. 101. Bethesda, MD: National Council on Radiation Protection and Measurements, 1989.

^aIncludes 10 mSv (1 rem) from high LET (α) radiation.

^bIncludes 0.5 mSv (50 mrem) from high LET (α) radiation.

LET, linear energy transfer.

average annual effective dose equivalents to airline crews, who are not regarded as radiation workers, ~ 1.7 mSv/year (170 mrem/year), typically exceed the annual effective dose equivalents to diagnostic radiology and nuclear medicine personnel.

Collective Effective Dose Equivalent

Another way to assess the impact on human health from population radiation exposures is to evaluate the average effective dose equivalent and the size of the exposed population. The product of these two factors is the *collective effective dose equivalent*, expressed in person-sieverts (person-Sv or person-rem). Table 23-4 shows an estimate of the U.S. population collective effective dose equivalent from natural background and technologic radiation sources.

Table 23-5 shows the collective effective dose equivalent for some occupationally exposed groups. Although uranium miners have some of the highest average individual exposures, they have among the lowest collective effective dose equivalent due to the relatively small size of the work force compared to other occupa-

TABLE 23-4. SUMMARY OF THE ANNUAL COLLECTIVE EFFECTIVE DOSE EQUIVALENT FROM ALL SOURCES IN THE UNITED STATES

	Person-Sv/yr	Person-rem/yr	% Total
Radon	460,000	46,000,000	~ 55
Natural background (excluding radon)	230,000	23,000,000	~ 28
Medical radiation	123,000	12,300,000	~ 15
Enhanced natural sources (consumer products, domestic water, etc.)	20,000	2,000,000	< 1
Occupational	2,000	200,000	$< < 1$
Miscellaneous environmental sources	160	16,000	$< < < 1$
Nuclear power	130	13,000	$< < < 1$
Rounded total	$\sim 835,000$	$\sim 83,500,000$	100

Adapted from National Council on Radiation Protection and Measurements. *Ionizing radiation exposure of the population of the United States*. NCRP report no. 93. Bethesda, MD: National Council on Radiation Protection and Measurements, 1987.

TABLE 23-5. SUMMARY OF THE ANNUAL COLLECTIVE EFFECTIVE DOSE EQUIVALENTS FOR WORKERS IN VARIOUS OCCUPATIONALLY EXPOSED CATEGORIES IN THE UNITED STATES

Occupational category	Annual collective effective dose equivalents	
	Person-Sv	Person-rem
Nuclear power operations	550	55,000
Medicine	420	42,000
Industry (other than nuclear fuel cycle)	390	39,000
Airline crews and attendants	170	17,000
Uranium miners	110	11,000

Adapted from National Council on Radiation Protection and Measurements. *Exposure of the U.S. population from occupational radiation*. NCRP report no. 101. Bethesda, MD: National Council on Radiation Protection and Measurements, 1989.

tions. The total annual collective effective dose equivalent for all occupationally exposed individuals (~1.6 million) is estimated to be $\sim 2.0 \times 10^3$ person-Sv (2.0×10^5 person-rem). The annual collective effective dose equivalent attributable to natural background radiation for this same population is $\sim 3.2 \times 10^3$ person-Sv (3.2×10^5 person-rem). Therefore, occupational exposure represents an increase of ~63% over that from natural background radiation for radiation workers.

In general, the collective effective dose equivalent has been decreasing in medicine and many other occupationally exposed groups. This decrease is due to both smaller occupationally exposed populations (e.g., nuclear fuel cycle workers) and improved health physics practices that have resulted in lower individual doses.

Genetically Significant Dose

The genetically significant equivalent dose (GSD) is used to assess the genetic risk or detriment to the whole population from a source of radiation exposure. The GSD (measured in sieverts or rem) is defined as that equivalent dose that, if received by every member of the population, would be expected to produce the same genetic injury to the population as do the actual doses received by the irradiated individuals. In determining the GSD, the equivalent dose to the gonads of each exposed individual is weighted for the number of progeny expected for a person of that sex and age.

The annual GSD from all radiation sources is ~1.3 mSv (130 mrem) (Table 23-6). The single largest contribution, 1.02 mSv (102 mrem), or 78%, is from natural background radiation, principally cosmic rays, terrestrial exposure, and radionuclides in the body. The contribution of radon to the GSD is minimal, ~100 μ Sv (10 mrem), because almost all of the dose from this source is received by the bronchial epithelium.

Technologic sources contribute to the GSD primarily in the form of medical exposures, ~200 μ Sv (20 mrem) or 15%, of which approximately one-third is attributable to irradiation of males and two-thirds to irradiation of females. The female component is higher because of the location of the ovaries within the pelvis, which places them in the primary beam during most abdominopelvic examinations. The genetically significant dose and genetically significant equivalent dose can be used interchangeably in reference to medical exposures insofar as they predominantly use low linear energy transfer (LET) radiation.

**TABLE 23-6. ANNUAL GENETICALLY SIGNIFICANT DOSE (GSD)
IN THE U.S. POPULATION**

Source	Contribution to the GSD		% of Total
	mSv	mrem	
Natural sources			
Radionuclides in the body	0.36	36	78
Terrestrial	0.28	28	
Cosmic	0.27	27	
Radon	0.10	10	
Cosmogenic	0.01	1	
Subtotal	~1.02	~102	
Technological sources			
Diagnostic x-rays	0.20	20	22
Nuclear medicine	0.02	2	
Miscellaneous (consumer products)	0.05	5	
Other (e.g., occupational; nuclear fuel cycle; fallout; misc environmental sources)	<0.01	<1	
Subtotal	~0.28	~28	
Total	1.30	130	100

Adapted from National Council on Radiation Protection and Measurements. *Ionizing radiation exposure of the population of the United States*. NCRP report no. 93. Bethesda, MD: National Council on Radiation Protection and Measurements, 1987.

Summary

The average annual effective dose equivalent to the U.S. population from all radiation sources is obtained by dividing the annual collective effective dose equivalent by the size of the U.S. population. The result is ~3.6 mSv/year (360 mrem/year) or ~10 μ Sv/day (1 mrem/day) for all people in the U.S. from all sources, exclusive of smoking.

The average annual effective dose equivalent is somewhat misleading for categories such as the medical use of radiation, in that many people in the population are neither occupationally nor medically exposed, yet are included in the average. Figure 23-1 presents a summary of the annual average effective dose equivalent for the U.S. population from the radiation sources previously discussed. This figure does not include smoking for reasons previously mentioned; however, if included [~2.8 mSv/year (~280 mrem/year)], smoking would be the single largest contributor (~45%) to the average population effective dose equivalent. It is clear from these comparisons that the lung is the organ receiving the highest dose equivalent for both smokers (polonium-210) and nonsmokers (radon-222).

23.2 PERSONNEL DOSIMETRY

Personnel radiation exposure must be monitored for both safety and regulatory purposes. This assessment may need to be made over a period of several minutes to several

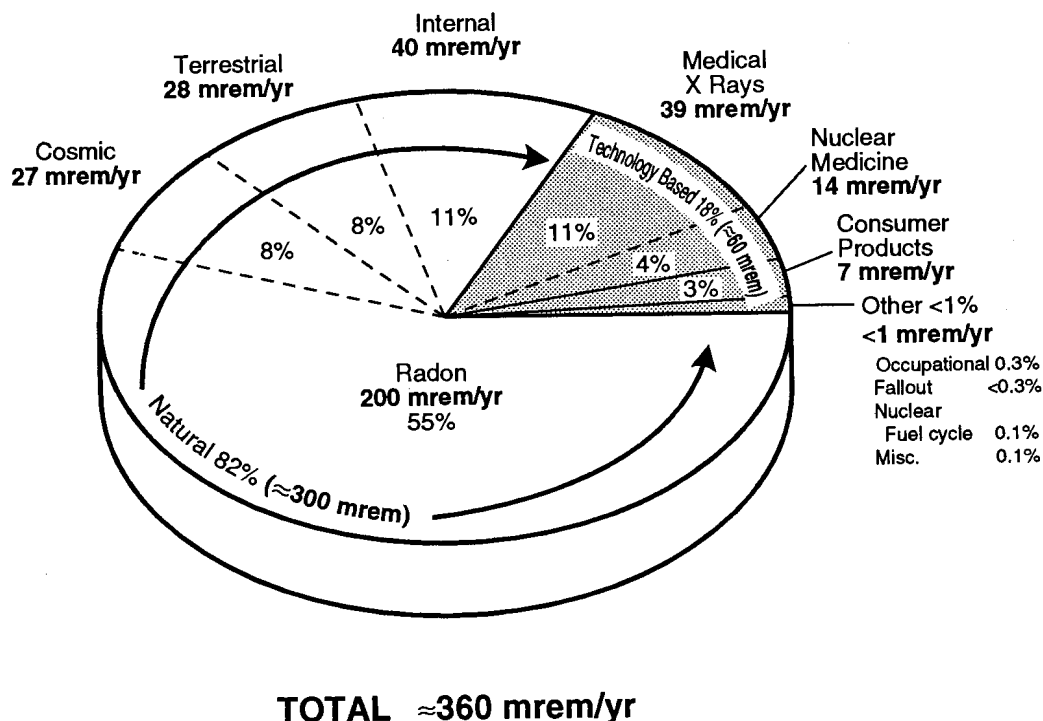


FIGURE 23-1. The percentage contribution of various radiation sources to the total average effective dose equivalent (360 mrem/year) in the U. S. population. Divide mrem/year by 100 to obtain mSv/year. (Adapted from National Council on Radiation Protection and Measurements. *Ionizing radiation exposure of the population of the U.S.* NCRP report no. 93. Bethesda, MD: National Council on Radiation Protection and Measurements, 1987.)

months. There are three main types of individual radiation recording devices called *personnel dosimeters* used in diagnostic radiology and nuclear medicine: (a) *film badges*, (b) *dosimeters using storage phosphors* (e.g., thermoluminescent dosimeters, TLDs), and (c) *pocket dosimeters*, each with its own specific advantages and disadvantages.

Ideally, one would like to have a single personnel dosimeter capable of meeting all of the dosimetry needs in medical imaging. The ideal dosimeter would instantaneously respond, distinguish between different types of radiation, and accurately measure the dose equivalent from all forms of ionizing radiation with energies from several keV to MeV, independent of the angle of incidence. In addition, the dosimeter would be small, lightweight, rugged, easy to use, inexpensive, and unaffected by changes in environmental conditions (e.g., heat, humidity, pressure) and nonionizing radiation sources. Unfortunately, no such dosimeter exists; however, the majority of these characteristics can be satisfied to some degree by selecting the dosimeter best suited for a particular application.

Film Badges

The film badge is the most widely used dosimeter in diagnostic radiology and nuclear medicine. It consists of a small sealed film packet (similar to dental x-ray film) placed inside a special plastic holder that can be clipped to clothing. Just as

with conventional x-ray film, radiation striking the emulsion causes a darkening of the developed film (see Chapter 6). The amount of darkening is measured with a densitometer and increases with the absorbed dose to the film emulsion. The film emulsion contains grains of silver bromide resulting in a higher effective atomic number than tissue; therefore, the dose to film is not equal to the dose to tissue. However, with the selective use of several metal filters over the film (typically lead, copper, and aluminum), the relative optical densities of the film underneath the metal filters can be used to identify the general energy range of the radiation and allow for the conversion of the film dose to tissue dose. Film badges typically have an area where the film is not covered by a metal filter or plastic and thus is directly exposed to the radiation. This “open window” is used to detect medium and high-energy beta radiation that would otherwise be attenuated (Fig. 23-2).

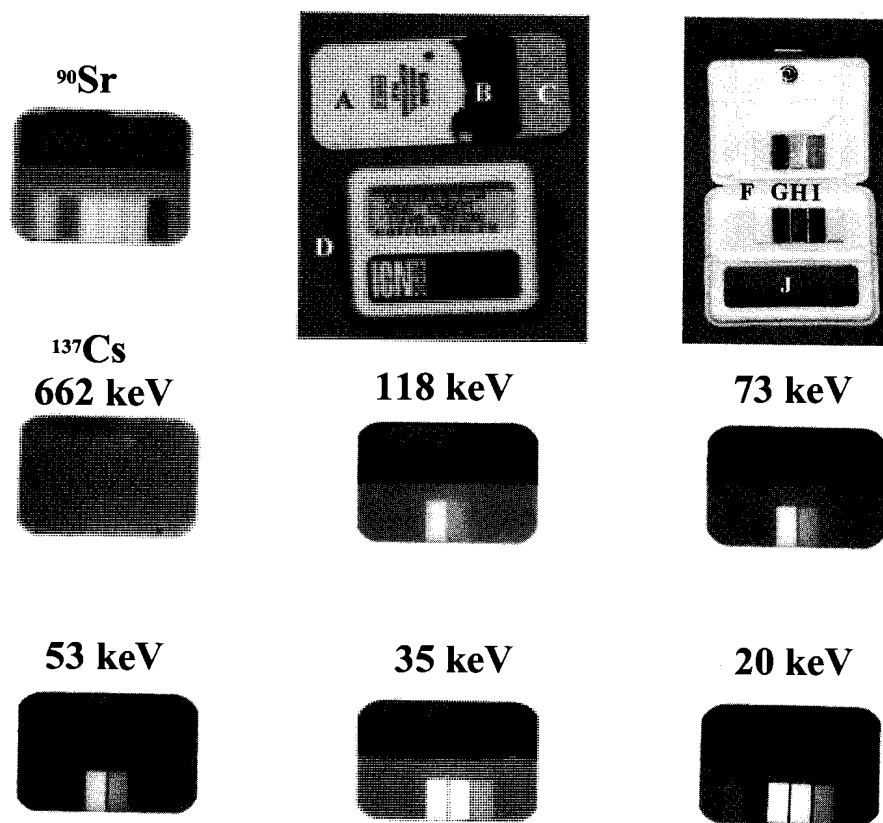


FIGURE 23-2. A film pack (A) consists of a black (light opaque) envelope (B) containing the film (C). The film pack is placed in the plastic film badge (D) sandwiched between two sets of Teflon (F), lead (G), copper (H), and aluminum (I) filters. Film badges typically have an area where the film pack is not covered by a filter or the plastic of the badge and thus is directly exposed to the radiation. This “open window” area (J) is used to detect medium and high-energy beta radiation that would otherwise be attenuated. The relative darkening on the developed film (filter patterns) provides a crude but useful assessment of the energy of the radiation exposure. The diagram shows typical filter patterns from exposure to a high energy beta emitter (Sr-90), a high energy gamma emitter (Cs-137), and x-rays with effective energies from 20 to 118 keV. Film badge and filter patterns courtesy of ICN Worldwide Dosimetry Service, Irvine, CA.

Most film badges can record doses from about 100 μGy to 15 Gy (10 mrad to 1,500 rad) for photons and from 500 μGy to 10 Gy (50 mrad to 1,000 rad) for beta radiation. The film in the badge is usually replaced monthly and sent to the commercial supplier for processing. An exposure report is received from the vendor in approximately 2 weeks. The developed film is usually kept by the vendor, providing a permanent record of radiation exposure. The dosimetry report lists the “shallow” dose, corresponding to the skin dose, and the “deep” dose, corresponding to penetrating radiations.

Film badges are small, lightweight, inexpensive, and easy to use. However, exposure to excessive moisture or heat will damage the film emulsion, making dose estimates difficult or impossible. The film badge is typically worn on the part of the body that is expected to receive the largest radiation exposure or is most sensitive to radiation damage. Most radiologists, x-ray technologists, and nuclear medicine technologists wear the film badge at waist or shirt-pocket level. During fluoroscopy, film badges are typically placed at collar level in front of the lead apron to measure the dose to the thyroid and lens of the eye because the majority of the body is shielded from exposure. Pregnant radiation workers typically wear a second badge at waist level (behind the lead apron, if used) to assess the fetal dose.

Thermoluminescent and Optically Stimulated Luminescent Dosimeters

Some dosimeters contain storage phosphors, in which a fraction of the electrons, raised to excited states by ionizing radiation, become trapped in excited states. When these trapped electrons are released, either by heating or by exposure to light, they fall to lower energy states with the emission of light.

Thermoluminescent dosimeters (TLDs), discussed in Chapter 20, are excellent personnel and environmental dosimeters; however, they are somewhat more expensive and are not as widely used in diagnostic radiology as film badges. The most commonly used TLD material for personnel dosimetry is lithium fluoride (LiF). LiF TLDs have a wide dose response range of 10 μSv to 10³ Sv (1 mrem to 10⁵ rem) and are reusable. Another advantage of LiF TLDs is that the effective atomic number is close to that of tissue; therefore, the dose to the dosimeter is close to the tissue dose over a wide energy range. TLDs do not provide a permanent record, because heating the chip to read the exposure removes the deposited energy. TLDs are routinely used in nuclear medicine as extremity dosimeters; a finger ring containing a chip of LiF is worn on the hand expected to receive the highest exposure during radiopharmaceutical preparation and administration. Figure 23-3 shows a finger ring dosimeter and LiF chip. TLDs are also the dosimeter of choice when longer dose assessment intervals (e.g., quarterly) are required.

Dosimeters using optically stimulated luminescence (OSL) have recently become commercially available as an alternative to TLDs. The principle of OSL is similar to that of TLDs, except that the light emission is stimulated by laser light instead of by heat. Crystalline aluminum oxide activated with carbon ($\text{Al}_2\text{O}_3\text{:C}$) is commonly used. Like TLDs, OSL dosimeters have a broad dose response capability, and are capable of detecting doses as low as 10 μSv (1 mrem). However, OSL dosimeters have certain advantages over TLDs in that they can be reread several times and an image of the filter pattern can be produced to differentiate between static (i.e., instantaneous) and dynamic (i.e., normal) exposure.

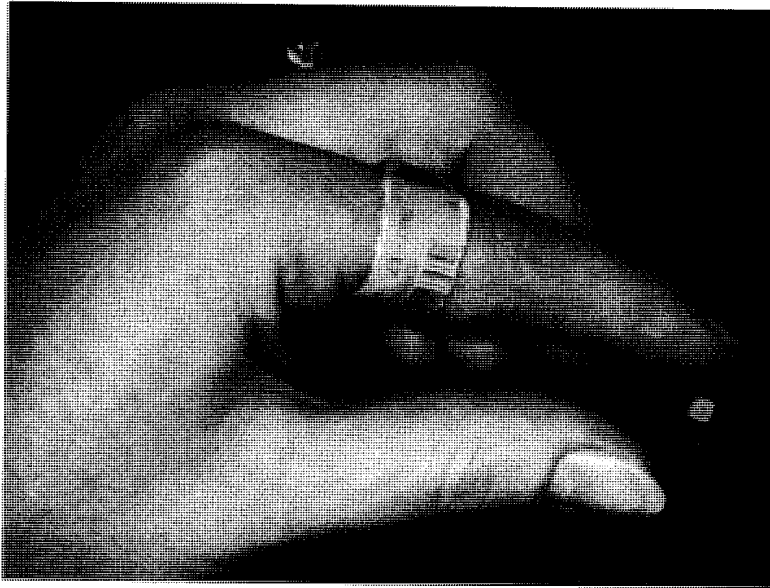


FIGURE 23-3. A small chip of LiF (*right*) is sealed in a ring (underneath the identification label) that is worn on the palmar surface such that the chip would be facing a radiation source held in the hand.

Pocket Dosimeters

The major disadvantage to film and TLD dosimeters is that the accumulated exposure is not immediately indicated. Pocket dosimeters measure radiation exposure, which can be read instantaneously. The analog version of the pocket dosimeter (pocket ion chamber) utilizes a quartz fiber suspended within an air-filled chamber on a wire frame, on which a positive electrical charge is placed. The fiber is bent away from the frame because of coulombic repulsion and is visible through an optical lens system upon which an exposure scale is superimposed. When radiation strikes the detector, ion pairs are produced in the air that partially neutralize the positive charge, thereby reducing coulombic repulsion and allowing the fiber to move closer to the wire frame. This movement is seen as a down-range excursion of the hairline fiber on the exposure scale (Fig. 23-4). Pocket ion chambers can typically detect photons of energies greater than 20 keV. The most commonly used models measure exposures from 0 to 200 mR or 0 to 5 R. These devices are small (the size of a pen) and easy to use; however, they may produce erroneous readings if bumped or dropped and, although reusable, do not provide a permanent record of exposure.

Exposure is defined as the amount of electrical charge produced by ionizing electromagnetic radiation per mass of air and expressed in units of coulombs per kg (C/kg) (see Chapter 3). The historical unit of exposure, the roentgen (R), is defined as $1 \text{ R} = 2.58 \times 10^{-4} \text{ C/kg}$. The dose to air can be calculated from the exposure whereby 1 R of exposure results in 8.76 mGy (0.876 rad) of air dose. The quantity exposure is still in common use in the United States and will be used exclusively in this chapter. However the related quantity of air kerma is predominantly used in many other countries.

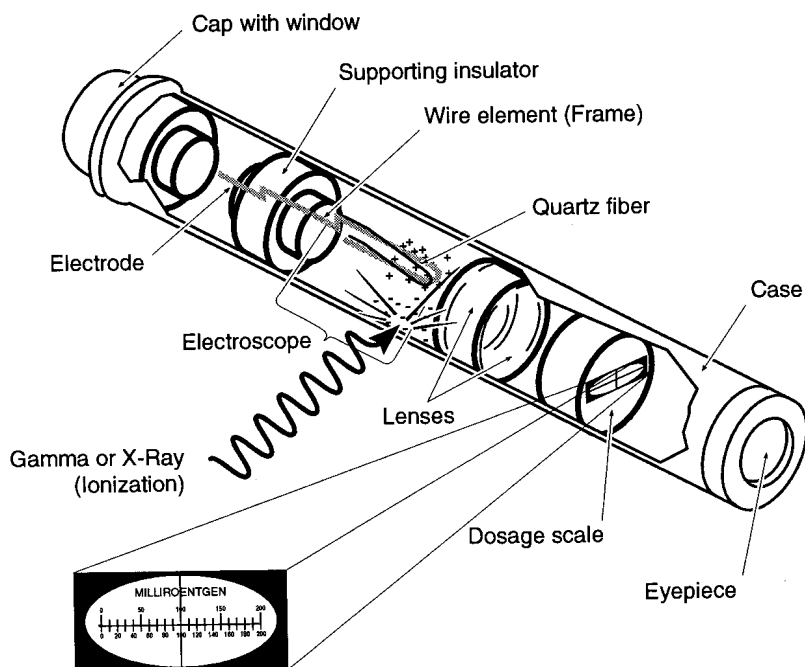


FIGURE 23-4. Cross section of an analog pocket ion chamber.

Digital pocket dosimeters can be used in place of the pocket ion chamber. These dosimeters use either Geiger-Mueller (GM) tubes or radiation-sensitive diodes and solid-state electronics to measure and display radiation dose in a range from approximately $10 \mu\text{Sv}$ (1 mrem) to 100 mSv (10 rem). Figure 23-5 shows analog and digital pocket dosimeters.



FIGURE 23-5. Analog and digital pocket dosimeters.

TABLE 23-7. SUMMARY OF PERSONNEL MONITORING METHODS

Method	Measures	Useful range		Permanent record	Uses and comments
Film badge	Beta; gamma and x-ray	(x and gamma) 0.1–15,000 mSv ^a (beta) 0.5–10,000 mSv ^a		Yes	Routine personnel monitoring; most common in diagnostic radiology and nuclear medicine
TLD	Beta; gamma and x-ray	0.01–10 ⁶ mSv ^a		No	Becoming more common but still more expensive than film; used for phantom and patient dosimetry
OSL	Beta; gamma and x-ray	0.01–10 ⁶ mSv ^a		No ^b	Advantage over TLD includes the ability to reread the dosimeters and distinguish between dynamic and static exposures
Pocket dosimeter	Gamma and x-ray	<i>Analog</i> 0 to 0.2 R 0 to 0.5 R 0 to 5 R	<i>Digital</i> 0–100 mSv ^a	No	Special monitoring (e.g., cardiac cath); permits direct (i.e., real-time) reading of exposure

^aMultiply mSv by 100 to obtain mrem.

^bOSL dosimeters are typically retained and can be reread by the manufacturer for approximately 1 year. OSL, optically stimulated luminance; TLD, thermoluminescent dosimeter.

Pocket dosimeters are utilized when high doses are expected, such as during cardiac catheterization or manipulation of large quantities of radioactivity. Table 23-7 summarizes the characteristics of the various personnel monitoring devices discussed above.

Problems with Dosimetry

Common problems associated with dosimetry include leaving dosimeters in a radiation field when not worn; radionuclide contamination of the dosimeter itself; lost and damaged dosimeters; and not wearing the dosimeter when working with radiation sources. If the dosimeter is positioned so that the body is between it and the radiation source, the attenuation will cause a significant underestimation of the true exposure. Most personnel do not remain in constant geometry with respect to the radiation sources they use. Consequently, the dosimeter measurements are generally representative of the individual's average exposure. For example, if the film badge is worn properly and the radiation field is multidirectional or the wearer's orientation toward it is random, then the mean exposure over a period of time will tend to be a good approximation ($\pm 10\%$ to 20%) of the individual's true exposure.

23.3 RADIATION DETECTION EQUIPMENT IN RADIATION SAFETY

A variety of portable radiation detection instruments are used in diagnostic radiology and nuclear medicine, the characteristics of which are optimized for specific applications. The portable GM survey meter and portable ionization chamber sat-

isfy the majority of the requirements for radiation protection in nuclear medicine. X-ray machine evaluations require specialized ion chambers capable of recording exposure, exposure rates, and exposure durations. All portable radiation detection instruments should be calibrated at least annually.

Geiger-Mueller Survey Instruments

GM survey instruments are used to detect the presence and provide a semiquantitative estimate of the magnitude of radiation fields. Measurements from GM counters typically are in units of counts per minute (cpm) rather than mR/hr, because the GM detector does not duplicate the conditions under which exposure is defined. In addition, the relationship between count rate and exposure rate with most GM probes is a complicated function of photon energy. However, with specialized energy compensated probes, GM survey instruments can provide approximate measurements of exposure rate (typically in mR/hr). The theory of operation of GM survey instruments was presented in Chapter 20.

The most common application of GM counters is as radioactive contamination survey instruments in nuclear medicine. A survey meter coupled to a thin window (~ 1.5 to 2 mg/cm^2), large surface area GM probe (called a "pancake" probe) is ideally suited for contamination surveys (see Fig. 20-7). Thin window probes can detect alpha ($>3\text{MeV}$), beta ($>45 \text{ keV}$), x-, and gamma ($>6 \text{ keV}$) radiations. These detectors are extremely sensitive to charged particulate radiations with sufficient energy to penetrate the window but are relatively insensitive to x- and gamma radiations. These detectors will easily detect natural background radiation (on the order of ~ 50 to 100 cpm at sea level). These instruments have long dead times resulting in significant count losses at high count rates. For example, a typical dead time of $100 \text{ } \mu\text{sec}$ will result in a $\sim 20\%$ loss at $100,000 \text{ cpm}$. Older portable GM survey instruments will saturate in high radiation fields and read zero, which, if unrecognized, could result in significant overexposures. Portable GM survey instruments are best suited for low-level contamination surveys and should not be used in high-radiation fields or when accurate measurements of exposure rate are required unless specialized energy compensated probes or other techniques are used to account for these inherent limitations.

Portable Ionization Chambers

Portable air-filled ionization chambers are used when accurate measurements of radiation exposure are required. These ionization chambers approximate the conditions under which the roentgen is defined (see Chapter 3). They have a wide variety of applications including measurements of x-ray machine outputs, assessment of radiation fields adjacent to brachytherapy or radionuclide therapy patients, and surveys of radioactive materials packages. The principles of operation of ion chambers are discussed in Chapter 20.

Air-filled ionization chamber measurements are influenced by changes in temperature, atmospheric pressure, photon energy, and exposure rate. However, these limitations are not very significant for conditions typically encountered in medical imaging. For example, a typical portable ion chamber survey meter will experience only $\sim 5\%$ loss for exposure rates approaching 10 R/hr ; specialized detectors can measure much higher exposure rates. Most portable ionization chambers respond

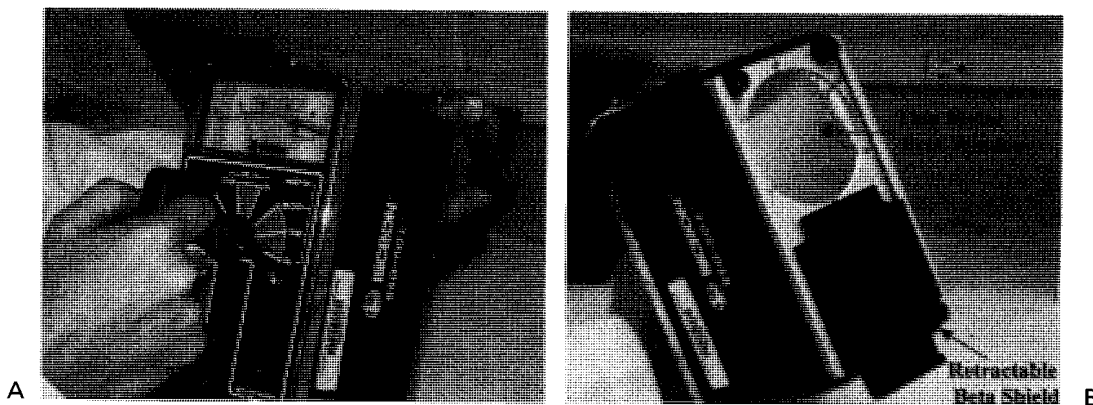


FIGURE 23-6. Portable ion chamber. **A:** An Sr-90 check source is used to test the instrument's response to radiation. **B:** The thick (1,000 mg/cm²) retractable plastic beta shield is in the open position exposing the thin (7 mg/cm²) aluminized Mylar beta window.

slowly to rapidly changing radiation exposure rates. These instruments must be allowed to warm up and stabilize before accurate measurements can be obtained. Some detectors are also affected by orientation and strong magnetic fields [e.g., from magnetic resonance imaging (MRI) scanners]. Ionization chambers are typically capable of measuring exposure rates between 1 mR/hr and 500 R/hr for photons greater than ~20 keV. Some ionization chambers have a cover over one end of the detector, which can be removed to assess the contribution of beta particles to the radiation field. In Fig. 23-6, the cover is retracted and a small radioactive check source is being used to verify the instrument's response to radiation. The slow response time and limited sensitivity of these detectors preclude their use as low-level contamination survey instruments or to locate a lost low-activity radiation source.

23.4 RADIATION PROTECTION AND EXPOSURE CONTROL

There are four principal methods by which radiation exposures to persons can be minimized: (a) time, (b) distance, (c) shielding, and (d) contamination control. Although these principles are widely used in all radiation protection programs, their application to diagnostic radiology and nuclear medicine will be addressed below.

Time

Although it is obvious that reducing the time spent near a radiation source will reduce one's radiation exposure, techniques to minimize time in a radiation field are not always recognized or practiced. To begin with, not all sources of radiation produce constant exposure rates. Diagnostic x-ray machines typically produce high exposure rates over brief time intervals. For example, a typical chest x-ray produces an entrance skin exposure of approximately 20 mR in less than $\frac{1}{20}$ of a second, an exposure rate of 1,440 R/hr. In this case, exposure is minimized by not activating the x-ray tube when staff are near the radiation source. Nuclear medicine proce-

TABLE 23-8. TYPICAL EXPOSURE RATE AT 1 M FROM ADULT NUCLEAR MEDICINE PATIENT AFTER RADIOPHARMACEUTICAL ADMINISTRATION

Study	Radiopharmaceutical	Activity (mCi) ^a	Exposure rate At 1 m (mR/hr)
Thyroid cancer therapy	I-131 (NaI)	200	~35
Tumor imaging	F-18 (FDG)	10	~5.0
Cardiac gated imaging	Tc-99m RBC	20	~0.9
Bone scan	Tc-99m MDP	25	~1.2
Tumor imaging	Ga-67 citrate	3	~0.4
Liver-spleen scan	Tc-99m sulfur colloid	4	~0.2
Myocardial perfusion imaging	Tl-201 chloride	3	~0.1

^aMultiply by 37 to obtain MBq. Exposure rate measurements courtesy of M. Hartman, Univ. California, Davis.

dures, however, typically produce lower exposure rates for extended periods of time. For example, the typical exposure rate at 1 m from a patient after the injection of ~740 MBq (20 mCi) of technetium 99m methylenediphosphonate (Tc-99m MDP) for a bone scan is ~1 mR/hr, which, through radioactive decay and urinary excretion, will be reduced to ~0.5 mR/hr at 2 hours after injection, when imaging typically begins. Therefore, knowledge of both the exposure rate and how it changes with time are important elements in minimizing personnel exposure. Table 23-8 shows the typical exposure rates at 1 m from adult nuclear medicine patients after the administration of commonly used radiopharmaceuticals.

The time spent near a radiation source can be minimized by having a thorough understanding of the tasks to be performed and the appropriate equipment to complete them in a safe and timely manner. For example, elution of a Mo-99/Tc-99m generator and subsequent radiopharmaceutical preparation requires several steps with a high activity source. It is important to have practiced these steps with a non-radioactive source to learn how to manipulate the source proficiently and use the dose measurement and preparation apparatus. In a similar fashion, radiation exposure to staff and patients can be reduced during fluoroscopy if the operator is well acquainted with the procedure to be performed.

Distance

The exposure rate from a point source of radiation (i.e., a source whose physical dimensions are much less than the distance from which it is being measured) decreases as the distance from the source squared. This *inverse square law* is the result of the geometric relationship between the surface area (A) and the radius (r) of a sphere: $A = 4\pi r^2$. Thus, if one considers an isotropic point radiation source at the center of the sphere, the surface area over which the radiation is distributed increases as the square of the distance from the source (i.e., the radius). If the exposure rate from a source is X_1 at distance d_1 , at another distance d_2 the exposure rate X_2 will be

$$X_2 = X_1(d_1/d_2)^2 \quad [23-1]$$

For example, if the exposure rate at 20 cm from a source is 90 mR/hr, doubling the distance will reduce the exposure by $(1/2)^2 = 1/4$ to 22.5 mR/hr; increasing the distance to 60 cm decreases the exposure by $(1/3)^2 = 1/9$ to 10 mR/hr (Fig. 23-7).

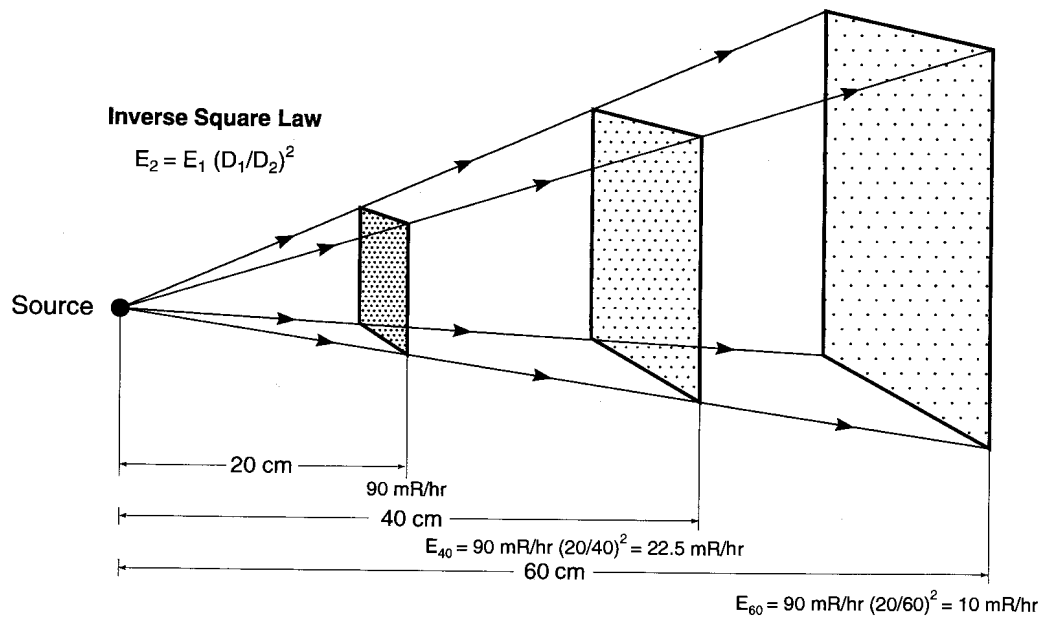


FIGURE 23-7. The inverse square law.

This relationship is only valid for point sources (i.e., sources whose dimensions are small with respect to the distances d_1 and d_2). Thus, this relationship would not be valid near (e.g., less than 1 m from) a patient injected with radioactive material. In this case, the exposure rate decreases less rapidly than $1/(\text{distance})^2$.

This rapid change in exposure rate with distance is familiar to all who have held their hand over a heat source, such as a stove top or candle flame. There is only a fairly narrow range over which the infrared (thermal, nonionizing) radiation felt is comfortable. A little closer and the heat is intolerable; a little further away and the heat is barely perceptible. Infrared radiation, like all electromagnetic radiation, follows the inverse square law.

Scattered radiation from a patient, x-ray tabletop, or shield is also a source of personnel radiation exposure. For diagnostic energy x-rays a good rule of thumb is that at 1 m from a patient at 90 degrees to the incident beam, the radiation intensity is ~0.1% to 0.15% (0.001 to 0.0015) of the intensity of the beam incident upon the patient for a ~400 cm² area (typical field area for projection imaging). As a practical matter, all personnel should stand as far away as possible during x-ray procedures without compromising the diagnostic quality of the exam. The NCRP recommends that personnel should stand at least 2 m from the x-ray tube and the patient and behind a shielded barrier or out of the room, whenever possible.

Because it is often not practical to shield the technologist from the radiation emitted from the patient during a nuclear medicine exam (see Shielding, below), distance is the primary dose reduction technique. The majority of the nuclear medicine technologist's annual whole-body radiation dose is received during patient imaging. Imaging rooms should be designed to maximize the distance between the imaging table and the computer terminal where the technologist spends the majority of the time during image acquisition and processing.

TABLE 23-9. EFFECT OF DISTANCE ON EXPOSURE WITH COMMON RADIONUCLIDES USED IN NUCLEAR MEDICINE

Radionuclide 370 MBq (10 mCi)	Exposure rate ^a (mR/hr) at 1 cm	Exposure rate ^a (mR/hr) at 10 cm (~4 inches)
Ga-67	7,500	75
Tc-99m	6,200	62
I-123	16,300	163
I-131	21,800	218
Xe-133	5,300	53
Tl-201	4,500	45

^aCalculated from the T_{20} .

Other than low activity sealed sources [e.g., a 370-kBq (10- μ Ci) Co 57 marker], unshielded radiation sources should never be manipulated by hand. The use of tongs or other handling devices to increase the distance between the source and the hand substantially reduces the exposure rate. Table 23-9 lists the exposure rates from radionuclides commonly used in nuclear medicine and the 100-fold exposure rate reduction achieved simply by increasing the distance from the source from 1 to 10 cm.

Shielding

Shielding is used in diagnostic radiology and nuclear medicine to reduce exposure to patients, staff, and the public. The decision to utilize shielding, and its type, thickness, and location for a particular application, are functions of the photon energy, number, intensity, and geometry of the radiation sources, exposure rate goals at various locations, and other factors. The principles of attenuation of x- and gamma rays are reviewed in Chapter 3.

Shielding in Diagnostic Radiology

The major objective of radiation shielding is to protect individuals working with or near x-ray machines from radiation emitted by the machines. Methods and technical information for the design of shielding for diagnostic x-ray rooms are found in NCRP report no. 49, entitled *Structural Shielding Design and Evaluation for Medical Use of X-rays and Gamma Rays of Energies up to 10 MeV*, published in 1976. The shielding methods in this report have served well for many years. However, the attenuation data in it for diagnostic x-ray equipment was measured using single-phase generators, whereas most diagnostic x-ray equipment today uses three-phase and high-frequency inverter generators. Furthermore, the work load data do not adequately describe current practice. For instance, the 400-speed film-screen image receptors used today require much less exposure than those used in the early 1970s. In addition, a general diagnostic room rarely uses a fixed kVp technique for all studies, but a range of kVps that on average are significantly lower and produce radiation that is much less penetrating. NCRP report no. 49 also incorporates many overly conservative assumptions, such as neglecting the attenuation of the primary beam by radiographic image receptors. The dose limits for members of the public have been reduced fivefold, to 100 mrem per year, since its publication. Strict

adherence to the methods in NCRP report no. 49 would result, in many cases, in unnecessary weight and expense in the shielding of x-ray rooms.

Many of the tables in this section are excerpted from this report, although the computational methods described herein do not exactly follow those in the report. A new shielding methodology is being prepared to replace NCRP report no. 49, with differences chiefly related to more accurate estimates of radiographic techniques in terms of work load (kVp and mAs), use of three-phase/high-frequency output and attenuation curves, consideration of primary beam attenuation by the radiographic screen-film detector, and changes in the radiation limits in adjacent areas. In particular, the new methodology recommends that occupational *restricted areas* be protected to radiation levels that are ten times *less* than currently specified (to 5 mSv/year as opposed to 50 mSv/year), and that occupancy factors of uncontrolled areas be adjusted to more realistic (lower) estimates than now specified.

Several factors must be considered when determining the amount and type of radiation shielding. An important concept is the ALARA principle—personnel radiation exposures must be kept “As Low As Reasonably Achievable” (see Regulatory Agencies and Radiation Exposure Limits, below). In addition, personnel exposures may not exceed limits established by regulatory agencies. In the U.S., these *annual* limits are 50 mSv (5 rem) for occupational workers and *50 times less* or 1 mSv (0.1 rem) for nonoccupational personnel (e.g., members of the public and nonradiation workers). Furthermore, the dose equivalent in an unrestricted area may not exceed 0.02 mSv (2 mrem) in any hour. When considering the shielding requirements for a room, it is convenient to use the *week* as the unit of time. Thus, in restricted (controlled) areas where only occupational workers are permitted, the maximal allowable exposure is 1 mSv (100 mrem) per week, although the current draft of the revision of NCRP report 49 recommends that controlled area levels be *reduced* by a factor of ten, to 0.1 mSv (10 mrem) per week. For unrestricted (uncontrolled) areas, the limit is 0.02 mSv (2 mrem) per week.

Sources of Exposure

The sources of exposure that must be shielded in a diagnostic x-ray room are *primary radiation*, *scattered radiation*, and *leakage radiation* (Fig. 23-8). Scatter and leakage radiation are together called *secondary* or *stray radiation*. Primary radiation, also called the *useful beam*, is the radiation passing through the open area defined by the collimator of the x-ray tube. The amount of primary radiation depends on the output of the x-ray tube (determined by the kVp, mR/mAs, and distance) per examination, the number of examinations performed during an average week, and the fraction of time the x-ray beam is directed toward any particular location. Scattered radiation arises from the interaction of the useful beam with the patient, causing a portion of the primary x-rays to be redirected. For radiation protection purposes scatter is considered as a separate radiation source with essentially the same photon energy spectrum (penetrability) as the primary beam. In general, the exposure from scattered radiation at 1 m from the patient is approximately 0.1% to 0.15% of the incident exposure to the patient for typical diagnostic x-ray energies and a 20 cm × 20 cm (400 cm²) field area. The scattered radiation is proportional to the field size, and can be calculated as a fraction of the reference field area. Leakage is the radiation that emanates from the x-ray tube housing other than the useful beam. Because leakage radiation passes through the shielding of the housing, its effective energy is very high (only the highest energy pho-

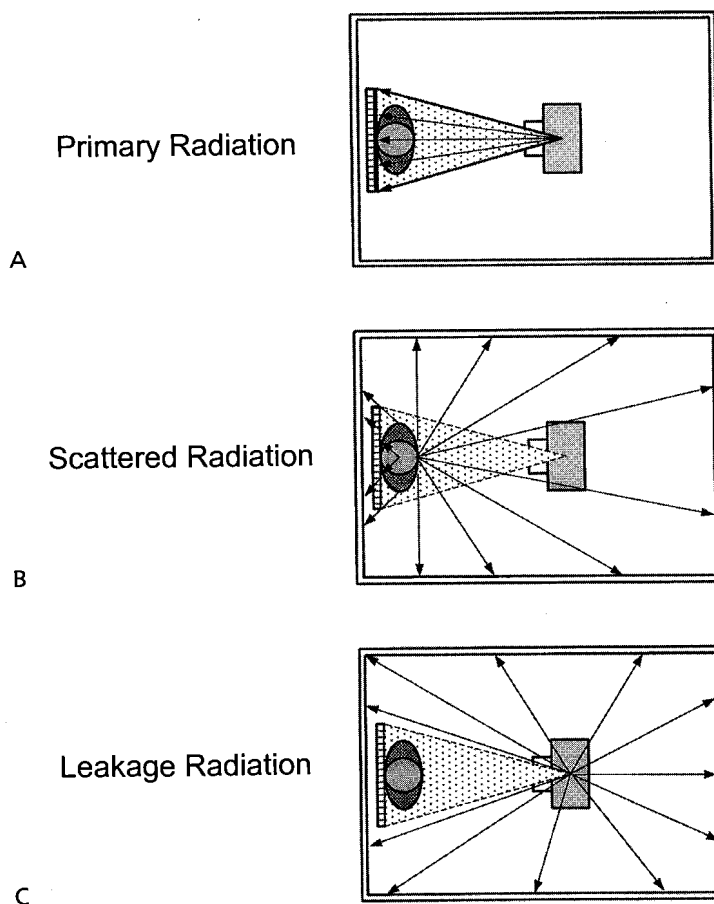


FIGURE 23-8. The various sources of exposure in a diagnostic x-ray room. **A:** Primary radiation emanating from the focal spot. **B:** Scattered radiation emanating from the patient. **C:** Leakage radiation emanating from the x-ray tube housing (other than the collimated primary radiation).

tons are transmitted). The exposure due to leakage radiation is limited by federal regulations to 100 mR/hr at 1 m from the tube housing when the x-ray tube is operated at the maximum allowable continuous tube current (usually 3 to 5 mA) at the maximum rated kVp, which is typically 150 kVp.

Factors Affecting Protective Barrier Specifications

X-ray shielding is accomplished by interposing a highly attenuating barrier, usually lead, between the source of radiation and the area to be protected in order to reduce the exposure to acceptable limits. The thickness of shielding needed to achieve the desired attenuation depends on the shielding material selected. Lead is the most commonly used material because of its high-attenuation properties and relatively low cost. Other shielding materials are also used, such as concrete, glass, leaded glass, leaded acrylic, and gypsum board.

The amount of attenuation necessary depends on five major factors: (a) The total *radiation exposure level*, X_T , represents the exposure (expressed in mR/week) at a given location in an adjacent area, which is a function of the technique (kVp), the work load, the calculated primary, scatter, and leakage radiation levels, and their

respective distances to the point in question, and is dependent on the subsequent factors listed below. (b) The *workload*, W , indicates the amount of x-rays that are produced per week. (c) The *use factor*, U , indicates the fraction of time during which the radiation under consideration is directed at a particular barrier. (d) The *occupancy factor*, T , indicates the fraction of time during a week that a single individual might spend in an adjacent area. (e) The *distance*, d , is measured from the sources of radiation to the area to be protected.

The workload, W , is expressed in units of mA·min/week, and is determined by the types of examinations and corresponding radiographic techniques (kV and mAs), detector characteristics (e.g., screen-film speed), average number of films per procedure, and average number of patients per week. A very busy radiographic room can have a workload of 1,000 mA·min/week, whereas rooms with light use (e.g., five or fewer patients per day) have a workload of 100 mA·min/week or less. For instance, a room with 20 patients per day, three films per patient, and 50 mAs per film has a workload of

$$W = 20 \text{ patients/day} \times 5 \text{ days/week} \times 3 \times \text{films/patient} \times 50 \text{ mAs/film} \times 1 \text{ min/60 sec} \\ = 250 \text{ mA}\cdot\text{min/week}.$$

When the workload cannot be determined exactly, it is prudent to reasonably overestimate the anticipated patient volume to provide sufficient shielding for adjacent areas. Table 23-10 lists suggested workload values for various types of radiographic rooms from NCRP report no. 49. These values are overestimated, as they are based on much slower (~100 speed) receptors compared to the 400 speed receptors now in common use. Higher kVp settings decrease the workload. This results from two factors: (a) the output of the x-ray tube (mR/mAs) increases as the kVp increases, and (b) less attenuation of the incident beam by the patient reduces the total mAs required to achieve a proper optical density on film.

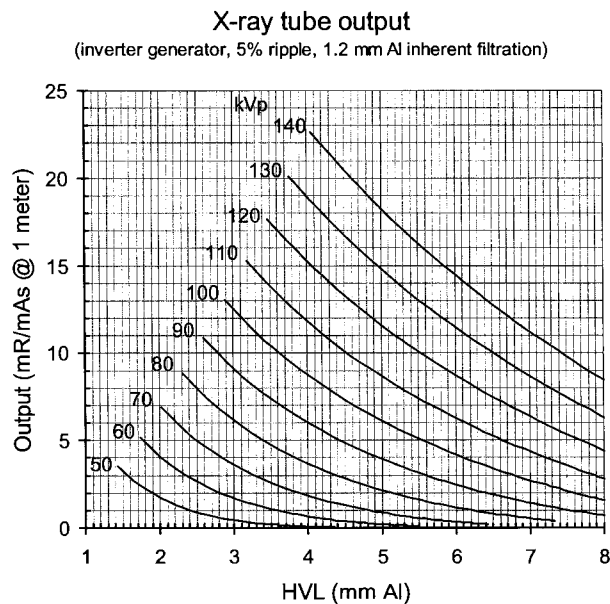
The x-ray tube output (mR/mAs at 1 m) is determined by direct measurement whenever possible. Alternatively, a table or a graph of the mR/mAs versus x-ray tube potential as a function of the half value layer (HVL) (Fig. 23-9A) or versus added filtration and HVL (Fig. 23-9B) can be used. For instance, an x-ray tube operated at 100 kVp with a 4-mm Al HVL delivers an output of 8.8 mR/mAs at 1 m as determined from the graph (Fig. 23-9A). Single-phase generators produce a lower output, and a reasonable estimate is determined by reducing the indicated output by a factor of about 1.3 (30%).

TABLE 23-10. SUGGESTED WORKLOAD VALUES FOR DIAGNOSTIC X-RAY INSTALLATIONS

Procedure	Daily patient load	W (mA·min/wk) (100 kVp)	W (mA·min/wk) (125 kVp)	W (mA·min/wk) (150 kVp)
Dedicated chest: 3 films/patient	60	250	150	100
Fluoroscopy and spot filming	24	750	300	150
General radiography	24	1000	400	200
Special procedures	8	700	280	140

Adapted from NCRP report no. 49.

A



B

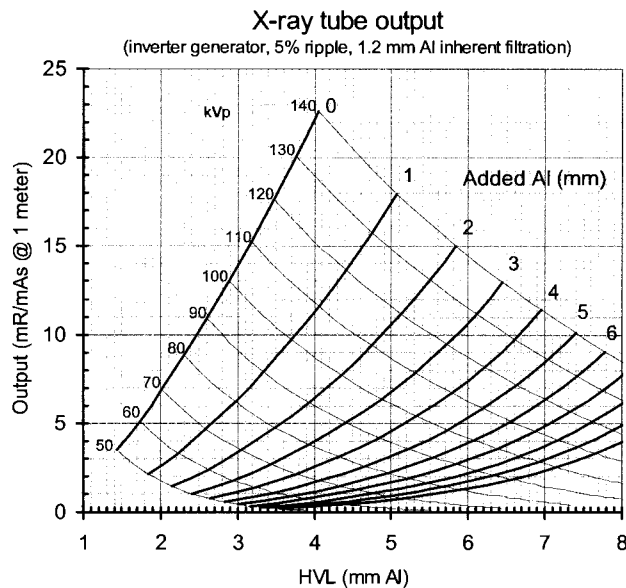


FIGURE 23-9. A: Tube output in mR/mAs at a distance of 1 m as a function of tube half value layer (HVL) in millimeters of aluminum and kVp for an inverter generator with 5% voltage ripple. Tube output is based on measured data and computer simulations from an x-ray tube with 1.2 mm inherent Al filtration. Each kVp curve has increasing added filtration from left to right. For single-phase generators, these curves overestimate the output by approximately 30% (i.e., divide the output value by 1.3). **B:** Tube output in mR/mAs as a function of HVL and added filtration. These curves indicate the amount of added filtration in the beam for a given HVL and kVp; each curve has increasing kVp from left to right. If the amount of added filtration and HVL are known, the tube output can be determined independent of the kVp. Additional filtration added to the beam increases the effective energy and decreases the output.

Exposure per week contributed by the primary beam, X_p , at 1 m from the source, is the product of W and the tube output measured at 1 m from the source:

$$X_p \text{ (mR/week)}_{1m} = W \text{ (mA} \cdot \text{min/week)} \cdot \text{tube output (mR/mA} \cdot \text{min)} \quad [23-2]$$

where the tube output is converted to mR/mA·min (e.g., 1 mR/mAs $\times$ 60 sec/min = 60 mR/mA·min).

Exposure contributed by scattered radiation from the patient is a fraction of the incident primary exposure. The scatter fraction at 1 m, S , is approximately 0.0015 (0.15%) at 1 m for a 400 cm² field at 125 kVp, but varies depending on kVp and the angle with respect to the primary beam (typically from 0.001 to 0.002). The scatter fraction also depends on field size (e.g., a 200-cm² field size produces about half the scatter of a 400-cm² field, whereas a 800-cm² area produces about twice the scatter).

Scattered radiation exposure is therefore the product of the incident exposure, the scatter fraction, S , and the ratio of the maximal field size relative to 400 cm². The incident exposure at 1 m, X_p , is corrected for distance to the scatterer, d_{sca} , as

$$X'_p = X_p / d_{sca}^2 \quad [23-3]$$

With these parameters accounted for, the scatter exposure per week, X_s , at a distance of 1 m from the scattering source (i.e., the surface of the patient) is calculated as

$$X_s \text{ (mR/week)}_{1m} = X'_p \times S \times [\text{field size (cm}^2\text{)} / 400 \text{ cm}^2] \quad [23-4]$$

Exposure due to leakage radiation cannot exceed 100 mR/hr at 1 m from the x-ray tube source for the maximum continuous allowable tube current, I (e.g., 3 to 5 mA), at the maximum kVp (e.g., 150 kVp). The maximal leakage radiation is therefore 100 mR/mA_{max} · hour at 1 m. This expression, converted to mA · min is 100 mR/(60mA_{max} · min) = 1.67 mR/(mA_{max} · min). The *total* leakage radiation per week, X_L , at 1 m from the tube housing is the product of the leakage radiation per mA·min and the weekly workload:

$$X_L \text{ (mR/week)}_{1m} = 1.67 \text{ mR} / (\text{mA}_{\text{max}} \cdot \text{min}) \times W \text{ (mA} \cdot \text{min/week)} \quad [23-5]$$

Note that a lower value of I (mA_{max}) increases the calculated contribution of leakage radiation for a given workload because less shielding in the tube housing is necessary to meet the leakage requirements.

Equations 23-2, 23-3, 23-4, and 23-5 provide the weekly exposure contributions of primary, scatter, and leakage radiation at a distance of 1 m from the respective radiation sources. The next steps in the shielding calculation must consider the x-ray system use (amount of time the primary x-rays are incident on a particular barrier) and the layout of the x-ray room (distance to the area to be protected and occupancy of the adjacent areas).

The use factor (U) of a wall is determined by the fraction of time radiation is incident on that wall during the week. A wall that intercepts the primary beam is called a *primary barrier* and is assigned a use factor according to typical room use. Values of U for primary radiation range from 0 to 1. For instance, a dedicated chest room always has the x-ray tube pointed toward the wall on which the chest bucky stand is mounted. This wall has a primary use factor of 1, while the other walls, floor, and ceiling of the room have primary use factors of 0. In a conventional radiographic room, the floor usually has the greatest primary beam use factor (e.g.,

0.75); other walls and ceiling have use factors that are estimated according to anticipated examinations and typical projections (e.g., 0.15 for wall B and 0.10 for wall C). *Secondary barriers always have a use factor of 1*, as scatter and leakage are incident on all surfaces whenever x-rays are produced. The walls, floor, and ceiling are considered secondary barriers. In those x-ray rooms with equipment having integrated primary barriers (e.g., a dedicated fluoroscopy suite where the image intensifier always intercepts the primary beam, a mammography room, and a CT room), there is no need to consider any room surfaces as primary barriers. The image detector (screen-film cassette) is not considered a primary barrier with existing (NCRP Report 49) recommendations.

The *occupancy factor, T*, is the estimate of the fraction of time that a *particular individual* will occupy an adjacent *uncontrolled* room during operation of the x-ray machine throughout the week. Table 23-11 lists suggested levels of occupancy. The occupancy factor modifies the radiation protection limit, X_{Limit} , of 2 mR/week as

$$X_{\text{Limit}} = \frac{2 \text{ mR/week}}{T} \quad [23-6]$$

Thus, an uncontrolled corridor with an occupancy factor of 0.25 has an exposure limit, X_{Limit} , of

$$X_{\text{Limit}} = \frac{2 \text{ mR/week}}{0.25} = 8 \text{ mR/week}$$

The distances from the sources of radiation to the point to be protected are a major factor in the determination of shielding thickness, because the inverse square law significantly affects the exposure levels. Four distances must be considered. The distance from the source to the scatterer, d_{sca} , is measured from the x-ray tube focal spot to the surface of the patient. This corrects the tube output (measured at 1 m) to the patient's skin (see Equation 23-3). The primary distance, d_{pri} , is measured from the x-ray tube focal spot to 0.3 m (1 foot) *beyond* the barrier to be shielded. The secondary radiation distance, d_{sec} , is measured from the closest surface of the patient (the scattering material) to 0.3 m beyond the barrier. The tube leakage distance, d_{leak} , is measured from the x-ray tube housing to 0.3 m beyond the barrier. In all cases, the shortest distance to the barrier should be used (e.g., for tabletop imaging, use the x-ray tube position that is closest to the barrier). Because the exposure levels X_p , X_s , and X_L are determined at 1 m, then division by d_{pri}^2 , d_{sec}^2 , and d_{leak}^2 , respectively, will yield the exposure levels at 0.3 m beyond the wall in the adjacent room.

TABLE 23-11. RECOMMENDED OCCUPANCY FACTOR (T)^a

Occupancy level	Type of area	T
Full	Work areas, offices, laboratories, nurses stations, living quarters, children's play area	1
Partial	Corridors, rest rooms, unattended parking lots	1/4
Occasional	Waiting rooms, toilets, stairways, elevators, closets	1/16

^aOutside areas used by pedestrians/traffic, as suggested by NCRP report no. 49.

Primary, scatter, and leakage radiation all contribute to the radiation levels in areas adjacent to an x-ray room. The total weekly exposure, X , *without shielding*, at a point 0.3 m beyond the wall in an area adjacent to an x-ray room is calculated as

$$X \text{ (exposure/week)} = \left[\left(\frac{X_p}{d_{\text{pri}}^2} \times U_{\text{pri}} \right) + \frac{X_s}{d_{\text{sec}}^2} + \frac{X_L}{d_{\text{leak}}^2} \right] \quad [23-7]$$

Shielding calculations determine the thickness of an attenuating material required to reduce X to acceptable levels. Knowledge of the HVL and tenth value layer (TVL) of the shielding material allows the necessary thickness to be calculated; the HVLs and TVLs of lead and concrete are listed in Table 23-12. For example, assume the calculated weekly exposure (unshielded) in an uncontrolled area is 40 mR, with an occupancy factor of 1. Reducing this exposure to 2 mR/week (i.e., 20-fold attenuation) requires one TVL (reduction by 10) and one HVL (reduction by two) of an attenuating material. For a 125-kVp beam with *lead* as the attenuator, this would approximately be 0.93 mm + 0.28 mm = 1.21 mm Pb; with *concrete*, the thickness required would be 66 mm + 20 mm = 86 mm concrete, where the TVL and HVL for lead and for concrete are from Table 23-12 for a 125-kVp beam.

An alternative method is to calculate the required attenuator thickness from the attenuation equation:

$$X_{\text{Limit}} = X e^{-\mu_{\text{eff}} T} \quad [23-8]$$

where T is the thickness of the attenuator and X_{Limit} is the maximum permissible exposure limit (e.g., 2 mR/week, *uncontrolled* area, 100 mR/week, *controlled* area). A key to the use of this equation is the determination of the *effective attenuation coefficient*, μ_{eff} , calculated from the HVL or TVL of the shielding material:

$$\mu_{\text{eff, HVL}} = 0.693/\text{HVL} \quad \text{and} \quad \mu_{\text{eff, TVL}} = 2.303/\text{TVL} \quad [23-9]$$

However, Equations 23-8 and 23-9 are exact only for a monoenergetic beam and narrow-beam geometry (Chapter 3). When applied to a polyenergetic spectrum, these equations are approximate. The more conservative approach is to calculate the effective attenuation coefficient using the TVL, which yields a smaller attenuation value than from the HVL. A lead attenuator at 125 kVp has $\mu_{\text{eff, HVL}} = 24.75 \text{ cm}^{-1}$ (2.475 mm^{-1}), and $\mu_{\text{eff, TVL}} = 21.43 \text{ cm}^{-1}$ (2.143 mm^{-1}). The TVL expression is the more conservative approach (greater transmission for a given attenuator thickness). Substituting the TVL expression into Equation 23-8 gives

$$X_{\text{Limit}} = X e^{-\frac{2.303}{\text{TVL}} T}$$

TABLE 23-12. HALF VALUE LAYER (HVL) AND TENTH VALUE LAYER (TVL) FOR LEAD AND CONCRETE AT DIAGNOSTIC X-RAY ENERGIES

Peak voltage (kV)	Lead (mm) HVL	Lead (mm) TVL	Concrete (mm) HVL	Concrete (mm) TVL
50	0.06	0.17	4.3	15
70	0.17	0.52	8.4	28
100	0.27	0.88	16	53
125	0.28	0.93	20	66
150	0.30	0.99	22.4	74

and solving for T yields the required attenuator thickness:

$$T = 0.434 \times \ln \left(\frac{X}{X_{\text{Limit}}} \right) \times \text{TVL} \quad [23-10]$$

where $\ln$ is the natural logarithm and T has the same units as the TVL. The thickness of *lead* required to reduce 40 mR to 2 mR for a 125-kVp beam ($\text{TVL}_{\text{Pb}} = 0.93$ mm) according to Equation 23-10 is

$$T = 0.434 \times \ln \left(\frac{40}{2} \right) \times 0.93 \text{ mm} = 1.209 \text{ mm}$$

which is very close to the sum of one TVL and one HVL of lead in the example described previously. If the occupancy factor in the room is reduced to 20%, then X_{Limit} becomes $(2 \text{ mR/week})/0.20 = 10 \text{ mR/week}$, and the thickness of lead is calculated as

$$T = 0.434 \times \ln \left(\frac{40}{10} \right) \times 0.93 \text{ mm} = 0.560 \text{ mm}$$

a reduction of $\sim 2.2\times$ in thickness. It is important to note that the TVL and HVL are functions of the kVp, and for a work load with a range of kVp techniques, a weighted kVp average approximates the beam characteristics for determining the HVL and TVL values.

Example of an X-Ray Room Shielding Calculation

The following example is intended to illustrate the shielding of one wall for a simple x-ray room, based on room configuration, radiation exposure limits, and estimated work load. Figure 23-10 shows a dedicated chest room and the use of adjacent areas. The work load is estimated to be 30 patients per day and two films per patient (posteroanterior and lateral projections), with a technique of 125 kVp and 5 mAs per film (a 400-speed screen-film system). The HVL of the beam is 4.0 mm Al, and the maximum continuous tube current at the maximum kVp is 5 mA. A high-frequency inverter generator provides the voltage to the x-ray tube.

The following steps are necessary to determine the shielding thickness required for the wall in question (in this example, Wall A):

1. The *weekly workload* (W) is calculated as

$$\begin{aligned} 30 \text{ patients/day} \times 2 \text{ films/patient} \times 5 \text{ mAs/film} \times 1/60 \text{ min/sec} \times 5 \text{ days/week} \\ = 25 \text{ mA}\cdot\text{min/week} \end{aligned}$$

2. The tube output for 125 kVp at 5 mm HVL (Fig. 23-9, interpolate between kVp curves) is $\sim 13 \text{ mR/mAs}$ at 1 m. This is equivalent to $780 \text{ mR/mA}\cdot\text{min}$ (multiply by 60 sec/min). The *primary exposure output per week at 1 m*, X_p , is determined using Equation 23-2:

$$X_p = 25 \text{ mA}\cdot\text{min/week} \times 780 \text{ mR/mA}\cdot\text{min} = 19,500 \text{ mR/week}$$

3. Incident exposure on the patient, X'_p , is determined using Equation 23-3. The source-to-object distance, d_{sca} , for chest imaging is approximately 1.5 m [assuming a 30 cm thick patient and a source-to-image distance (SID) of 180 cm], giving

$$X'_p = X_p/(d_{\text{sca}})^2 = 19,500/(1.5)^2 \approx 8,667 \text{ mR/week}$$

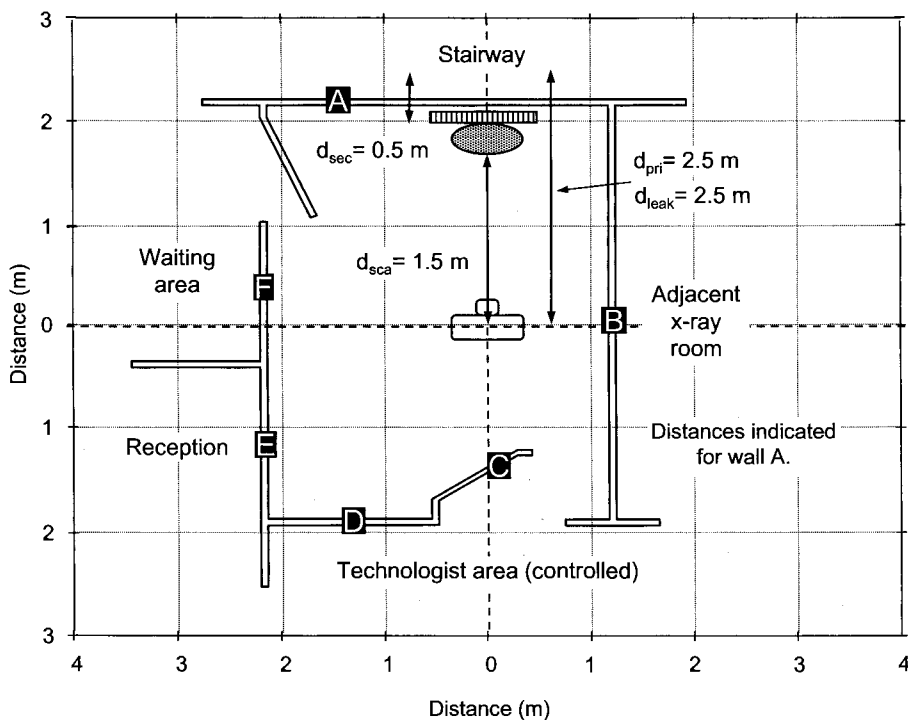


FIGURE 23-10. A dedicated chest room layout. Architectural schematics provide information regarding the x-ray tube position, detector locations, distance from each source of radiation to 0.3 m (1 foot) beyond each of the protective barriers in question, and the adjacent room usage (e.g., stairway versus waiting room versus dressing room). The overlay matrix indicates distance in meters. Distances for wall A from each of the radiation sources considered in shielding calculations are indicated.

A typical 35×43 cm field for chest imaging represents an area of $\sim 1,500$ cm². The scatter fraction at 90 degrees is 0.0015 for 125 kVp and a 400-cm² field area. The *scatter exposure output per week*, X_S , at 1 m from the patient is calculated using Equation 23-4:

$$X_S = 8,667 \times 0.0015 \times (1,500/400) \approx 49 \text{ mR/ week}$$

4. The *leakage exposure output* (X_L) at 1 m from the x-ray source is calculated using Equation 23-5 (from the workload (W) and the maximum continuous current, $I = 5$ mA):

$$X_L = 1.67 \times 25/5 \approx 8.4 \text{ mR/ week}$$

5. An occupancy factor (T) of 1/16 (0.0625) is assigned for the stairway (Table 23-11), and the permissible radiation exposure limit in this uncontrolled area is (Equation 23-6):

$$X_{\text{Limit}} = \frac{2 \text{ mR/week}}{0.0625} = 32 \text{ mR/week}$$

6. The *combined exposure without shielding* to a position 0.3 m (1 foot) beyond the protective barrier in question is determined from each of the sources of exposure

and the corresponding *closest* distances. Wall A for a dedicated chest room has a primary use factor (U_{pri}) of 1.0. The distances measured (Fig. 23-10) to 0.3 m beyond wall A are $d_{\text{pri}} = 2.5$ m, $d_{\text{leak}} = 2.5$ m, and $d_{\text{sec}} = 0.5$ m. The *total unshielded* exposure is calculated using Equation 23-7:

$$X = \left[\frac{19,500}{2.5^2} \times 1 + \frac{49}{0.5^2} + \frac{8.4}{2.5^2} \right] \cong 3,317 \text{ mR/week}$$

7. The lead barrier thickness required to satisfy the shielding requirements is calculated using Equation 23-10, where $\text{TVL}_{\text{Pb}}(125 \text{ kVp}) = 0.93$ mm:

$$T = 0.434 \times \ln \left(\frac{3,317}{32} \right) \times 0.93 \cong 1.87 \text{ mm}$$

This would correspond to 5 lb/ft² lead for wall A (Table 23-13). These seven steps are then accomplished for *each* wall, the floor, and the ceiling of the room. For the other barriers (walls, floor, and ceiling), the primary use factor is 0, and the required shielding will be much less.

The wall in this example is probably overshielded, because the calculations have ignored the attenuation of the screen-film cassette, the Bucky tray holder, and the patient. A measurement of the primary beam attenuation by the chest stand with cassette at 100 kVp indicates an equivalent attenuation of ~0.85 mm lead, so the required shielding with this consideration is about 1 mm lead, or 2½ lb/ft². In this case, it can be argued that 5 lb/ft² lead is excessive, as unnecessary expense and labor are required to meet unrealistic recommendations. Considerations like this are part of the rationale for the revision of NCRP report no. 49.

It is prudent to provide for the possibility of increased work loads in the future, since the installation of additional shielding is much more expensive than the cost of thicker materials for the original construction. Lead shielding is usually specified in architectural plans as weight per square foot (e.g., lb/ft²). Table 23-13 lists the lead thickness in millimeters and the corresponding weights in lb/ft². In addition, when shielding calculations indicate that a lead thickness *less than* 0.8 mm (2 lb/ft²) would be adequate (e.g., 1 lb/ft²), greater difficulty in handling and higher expense of the thinner lead material suggests that 2 lb/ft² lead

TABLE 23-13. THICKNESS AND WEIGHT EQUIVALENTS FOR LEAD SHEET

Thickness (inches)	Equivalent thickness (mm)	Nominal weight (lb/ft ²)	Actual weight (lb/ft ²)
1/64	0.40	1	0.92
3/128	0.60	1½	1.38
1/32	0.79	2	1.85
5/128	1.00	2½	2.31
3/64	1.19	3	2.76
7/128	1.39	3½	3.22
1/16	1.58	4	3.69
5/64	1.98	5	4.6
3/32	2.38	6	5.53
1/8	3.17	8	7.38

From NCRP report no. 49.

be the minimal thickness recommended. In the previous shielding calculation example, the other walls (as secondary barriers) require little extra shielding; a typical recommendation would be to install 2 lb/ft² lead shielding on these walls for the reasons stated above. On all walls, the shielding must extend from the base of the floor to a height of 7 feet. Floors and ceilings must also be considered in multiple floor buildings, using the same rationale and steps as outlined for the walls. Existing materials (e.g., concrete floors) often reduce or eliminate the need for additional shielding.

There are many alternative materials for shielding, including plate glass, leaded glass, leaded acrylic, gypsum drywall, plaster, barium plaster, and steel. Knowledge of their attenuation properties (HVLs and TVLs) permits their use instead of lead or concrete. For example, in screen-film mammography, where the tube voltage is rarely above 35 kVp and the scattered radiation is minimal, gypsum drywall is an appropriate shielding material. In most of these situations, shielding calculations indicate that two sheets of gypsum drywall (1.6 cm thickness) provide the necessary protection.

Computed Tomography Scanner Shielding

For a CT scanner, all walls in the room are secondary barriers, since the detector array provides the primary radiation barrier. The CT scanner presents a special situation requiring measured data from a typical scan (usually obtained from the manufacturer) to determine the amount of scattered radiation emanating from the gantry. Scattered radiation data are usually provided as exposure lines from the isocenter of the gantry on a per slice basis for known mAs (Fig. 23-11). In planning the shielding for the CT scanner, workload estimates are calculated from the average number of patients per week, the fraction of head versus body scans (scatter is markedly different for the head versus body), and the average mAs per patient (based on the number of slices and technique, e.g., conventional versus helical acquisition). Ignoring head scans (for the sake of simplicity) a sample workload might consist of 15 patients per day undergoing an abdominal CT scan (matching the scatter characteristics in Fig. 23-11). Of the total, assume 50% of the exams are pre- and postcontrast studies, and 50% are without contrast (an average of 1.5 studies per patient), and each exam consists of 40 slices acquired at 250 mAs per slice. The average number of slices per week is therefore

$$15 \text{ patients/day} \times 5 \text{ day/wk} \times 40 \text{ slices/study} \times 1.5 \text{ studies/patient} = 4,500 \text{ slices.}$$

Since the level of scattered radiation per slice is determined from the drawings, an overlay of the scatter distribution charts with the room layout is necessary. Consider the wall behind the CT gantry to be located 3.5 m from the gantry isocenter with an occupancy factor of 0.25. In this situation, the 0.075 mR/slice isoexposure contour crosses at a point approximately 0.3 m on the other side of the barrier. Thus, the total *unshielded* weekly exposure is calculated as

$$4,500 \text{ slices} \times \frac{0.075 \text{ mR}}{100 \text{ mAs}} \times \frac{250 \text{ mAs}}{\text{slice}} \cong 844 \text{ mR}$$

The occupancy factor, $T = 0.25$, increases the maximum allowable exposure in the adjacent area to

$$X_{\text{Limit}} = (2 \text{ mR/week})/0.25 = 8 \text{ mR/week}$$

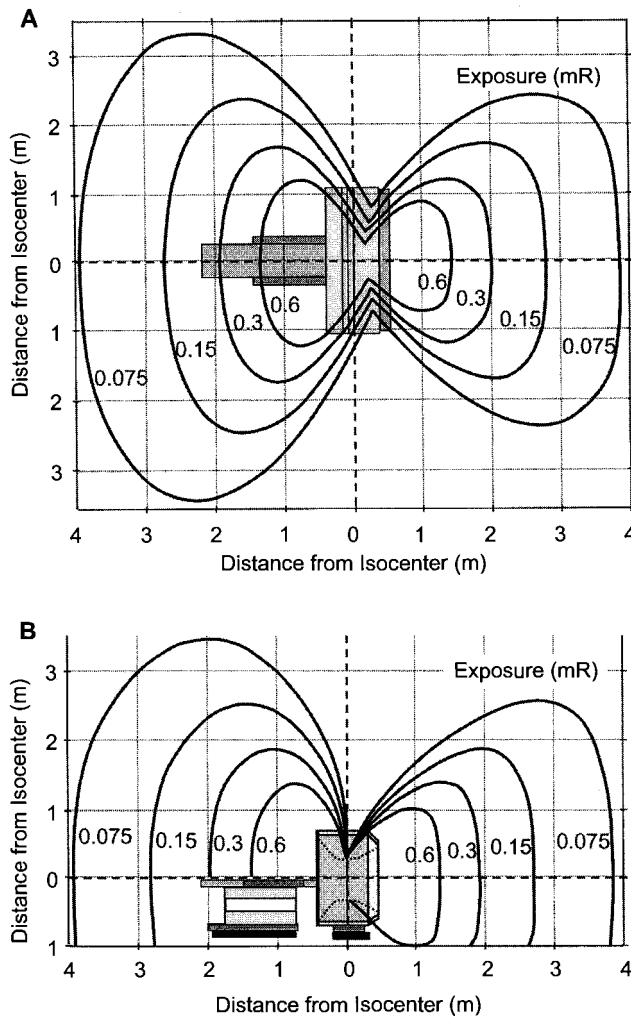


FIGURE 23-11. A multislice computed tomography (CT) scanner (25 mm collimator aperture) produces scatter and leakage radiation as a function of distance from the isocenter. Exposures are for 100 mAs per slice. Values must be scaled to the average mAs per slice and the total number of slices per week to determine the unshielded radiation exposure levels. Top view **(A)** of isoexposure contours on the isocenter plane extend horizontally from the gantry. Side view **(B)** of isoexposure contours on the isocenter plane extend vertically from the gantry.

The shielding required to reduce the radiation exposure to 8 mR per week is calculated using Equation 23-10 (using TVL_{Pb} for 125 kVp) as

$$T = 0.434 \times \ln \left(\frac{844}{8} \right) \times 0.93 = 1.88 \text{ mm}$$

Therefore, for this wall, 5 lb/ft² lead is required for that wall. A similar calculation for the other walls, floors and ceilings is also necessary. Elevational plots of scatter are used for protection requirements above and below the isocenter (Fig. 23-11B). Scattered radiation for CT scanners has increased significantly with the advent of multislice detector arrays and larger collimator openings. Shielding requirements for advanced CT systems have increased concurrently.

Personnel Protection in Diagnostic Radiology

Several items provide protection during a fluoroscopic/radiographic imaging procedure. The first and foremost is the *protective apron* worn by all individuals who must work in the room when the x-ray tube is operated. Lead equivalent thicknesses from 0.25 to 0.50 mm are typical. Usually, the lead is in the form of a rubber material to provide flexibility and handling ease. Aprons protect the torso of the body and are available in front or wrap-around designs, the latter being important when the back is exposed to the scattered radiation for a considerable portion of the time. Greater than 90% of the scattered radiation incident on the apron is attenuated by the 0.25 mm thickness at standard x-ray energies (e.g., tube voltage less than 100 kVp). Thicknesses of 0.35 and 0.50 mm lead equivalent in an apron give greater protection (up to 95% to 99% attenuation), but weigh 50% to 100% more than the 0.25 mm thickness. For long fluoroscopic procedures, the weight of the apron often becomes a limiting factor in the ability of the radiologist and the attending staff to complete the case without substitutions. Some apron designs, such as skirt-vest combinations, place much of the weight on the hips instead of the shoulders. The areas not covered by the apron include the arms, lower legs, the head and neck, and the back (except for wrap-around-type aprons).

Aprons also do not protect the thyroid gland or the eyes. Accordingly, there are *thyroid shields* and *leaded glasses* that can be worn by the personnel in the room. The leaded thyroid shield wraps around the neck to provide attenuation similar to that of a lead apron. Leaded glasses attenuate the incident x-rays to a lesser extent, typically 30% to 70%, depending on the content (weight) of the lead. Unfortunately, their weight is a major drawback. Normal, everyday glasses provide only limited protection, typically much less than 20% attenuation. Whenever the hands must be near the primary beam, *protective gloves* of 0.5-mm thick lead (or greater) should be worn.

In high work load interventional and cardiac catheterization laboratories, *ceiling-mounted or portable radiation protection barriers* are often used. These devices are placed between the patient and the personnel in the room. The ceiling mounted system is counterbalanced and easily positioned. Leaded glass or leaded acrylic shields are transparent and often provide greater attenuation than the lead apron.

Only persons whose presence is necessary should be in the imaging room during the exposure. All such persons must be protected with lead aprons or portable shields. In addition, no person should routinely hold patients during diagnostic examinations, certainly not those who are pregnant or under the age of 18 years. Mechanical supporting or restraining devices must be available and used whenever possible. In no instance shall the holder's body be in the useful beam, and should be as far away from the primary beam as possible.

Shielding in Nuclear Medicine

Radiation exposure rates in the nuclear medicine laboratory can range from over 100 R/hr [e.g., contact exposure from an unshielded ~37 GBq (1 Ci) Tc 99m generator eluate or a therapeutic dose ~11 GBq (300 mCi) of I 131] to natural background. The exposure rate at any distance from a particular radionuclide can be calculated using the *specific exposure rate constant* (Γ). The specific exposure rate

constant (expressed in units of $R\text{-cm}^2/\text{mCi}\cdot\text{hr}$) is the exposure rate in R/hr at 1 cm from 1 mCi of the specified radionuclide:

$$\text{Exposure rate (R/hr)} = \Gamma A/d^2 \quad [23-11]$$

where

Γ = specific exposure rate constant $R\text{-cm}^2/\text{mCi}\cdot\text{hr}$,

A = activity in mCi, and

d = distance in centimeters from a point source of radioactivity.

Because very low energy photons are significantly attenuated in air and other intervening materials, specific exposure rate constants usually ignore photons below a particular energy. For example, Γ_{20} represents the specific exposure rate constant for photons ≥ 20 keV.

Example 23-1: An unshielded vial containing 100 mCi of Tc 99m pertechnetate is left on a lab bench. For Tc 99m, $\Gamma_{20} = 0.6 R\text{-cm}^2/\text{mCi}\cdot\text{hr}$. What would be the exposure to a technician standing 50 cm away for 30 minutes?

Solution:

$$\text{Exposure} = (0.6 R\text{-cm}^2/\text{mCi}\cdot\text{hr}) (100 \text{ mCi}) (0.5 \text{ hr})/50^2 \text{ cm}^2 = 0.012 R = 12 \text{ mR}$$

Tungsten, lead, or leaded glass shields are used in nuclear medicine to reduce the radiation exposure from vials and syringes containing radioactive material. Table 23-14 shows specific exposure rate constants and lead HVLs for radionuclides commonly used in nuclear medicine. Syringe shields (Fig. 23-12) are used to reduce personnel exposure from syringes containing radioactivity during dose preparation and administration to patients. Syringe shields can reduce hand exposure from Tc-99m by as much as 100-fold.

Leaded glass shields are used in conjunction with solid lead shields in radiopharmaceutical preparation areas. Radiopharmaceuticals are withdrawn from vials surrounded by thick lead containers (called “lead pigs”) into shielded syringes behind the leaded glass shield in the dose preparation area (Fig. 23-13). Persons handling radionuclides should wear laboratory coats, disposable gloves, finger ring TLD dosimeters, and body dosimeters. The lead aprons utilized in diagnostic radiology are of limited value in nuclear medicine because, in contrast to their effectiveness in reducing exposure from low-energy scattered x-rays, they do not attenuate enough of the medium-energy photons emitted by Tc-99m (140 keV) to be practical (Table 23-15). Radioactive waste storage areas are shielded to minimize exposure rates. Mirrors mounted on the back wall of high-level radioactive material storage areas are often used to allow retrieval and manipulation of sources without direct exposure to the head and neck. Beta radiation is best shielded by low atomic number (Z) material (e.g., plastic or glass), which provides significant attenuation while minimizing bremsstrahlung x-ray production. For high-energy beta emitters (e.g., P-32), the low Z shield can be further shielded by lead to attenuate bremsstrahlung. With the exception of positron emission tomography (PET) facilities, which often shield surrounding areas from the 511-keV annihilation radiation, most nuclear medicine laboratories do not find it necessary to shield the walls within or surrounding the department.

TABLE 23-14. EXPOSURE RATE CONSTANTS (Γ_{20} AND Γ_{30})^a AND HALF VALUE LAYERS (HVL) OF LEAD FOR RADIONUCLIDES OF INTEREST TO NUCLEAR MEDICINE

Radionuclide	Γ_{20} (R-cm ² /mCi-hr) ^b	Γ_{30} (R-cm ² /mCi-hr) ^b	Half value layer in Pb (cm) ^{c,d}
Co-57	0.56	0.56	0.02
Co-60	12.87	12.87	1.2
Cr-51	0.18	0.18	0.17
Cs-137/Ba-137m	3.25	3.25	0.55
C-11, N-13, O-15	5.85	5.85	0.39
F-18	5.66	5.66	0.39
Ga-67	0.75	0.75	0.1
I-123	1.63	0.86	0.04
I-125	1.47	0.26	0.002
I-131	2.18	2.15	0.3
In-111	3.20	2.0	0.1
Ir-192	4.61	4.61	0.60
Mo-99/Tc-99m ^e	1.47	1.43	0.7
Tc-99m	0.62	0.60	0.03
Tl-201	0.45	0.45	0.02
Xe-133	0.53	0.53	0.02

^a Γ_{20} and Γ_{30} calculated from the absorption coefficients of Hubbell and Seltzer (1995) and the decay data table of Kocher (1981):

Hubbell JH, Seltzer SM. Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z = 1 to 92 and 48 additional substances of dosimetric interest. NISTIR 5632, National Institute of Standards and Technology, May 1995.

Kocher DC. Radioactive decay data tables. DOE/TIC-11026, Technical Information Center, U.S. Department of Energy, 1981.

^bMultiply by 27.03 to obtain $\mu\text{Gy}\cdot\text{cm}^2/\text{GBq}\cdot\text{hr}$ at 1 m.

^cThe first HVL will be significantly smaller than subsequent HVLs for those radionuclides with multiple photon emissions at significantly different energies (e.g., Ga-67) because the lower energy photons will be preferentially attenuated in the first HVL.

^dSome values were adapted from Goodwin PN. Radiation safety for patients and personnel. In: Freeman and Johnson's clinical radionuclide imaging, 3rd ed. Philadelphia: WB Saunders, 1984:370. Other values were calculated by the authors.

^eIn equilibrium with Tc-99m.

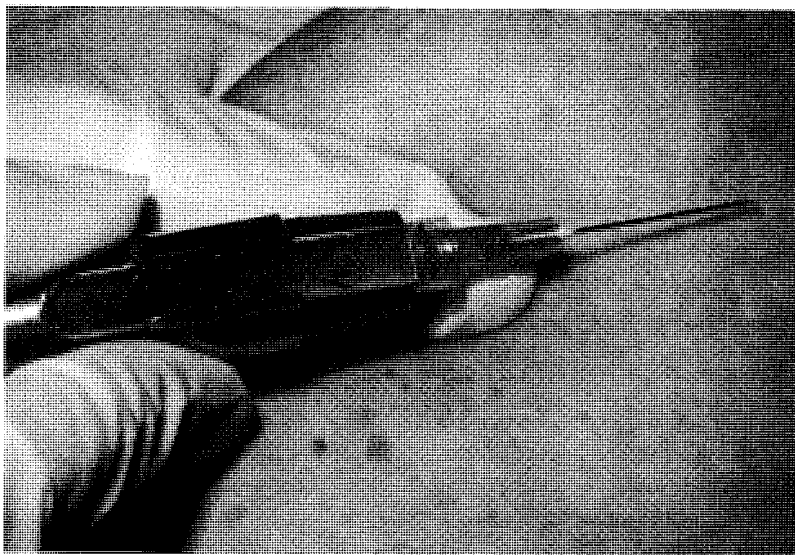


FIGURE 23-12. Syringe shield for a 3-cc syringe. The barrel is made of high-Z material (e.g., lead or tungsten) in which a leaded glass window is inserted so that the syringe graduations and the dose can be seen.

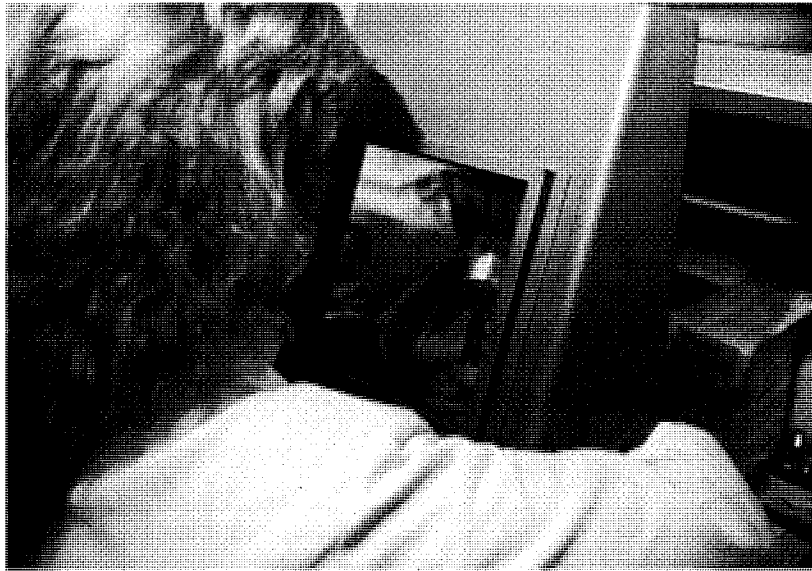


FIGURE 23-13. Dose preparation workstation. The technologist is drawing the radiopharmaceutical from a vial shielded by a "lead pig" into a syringe contained in a syringe shield. Further protection from radiation exposure is afforded by working behind the lead "L-shield" with a leaded glass window. The technologist is wearing a lab coat and disposable gloves to prevent contamination. A film badge is worn on the lab coat to record whole-body exposure and a thermoluminescent dosimeters (TLD) finger ring dosimeter is worn inside the glove to record the extremity exposure.

TABLE 23-15. EXPOSURE REDUCTION ACHIEVED BY A 0.5-mm LEAD EQUIVALENT APRON FOR VARIOUS RADIATION SOURCES

Source	Energy (keV)	Exposure reduction with 1 apron	No. of aprons to reduce exposure by 90% (i.e., 1 TVL)	Weight (lbs)
Scattered x-rays	10–30	>90%	~1	~15
Tc-99m	140	~70%	~2	~30
Cs-137	662	~6%	~36	~540

Protection of the Patient in Medical X-Ray Imaging

Tube Voltage and Beam Filtration

An important goal in diagnostic imaging is to achieve an optimal balance between image quality and dose to the patient. *Increasing the kVp* will result in a greater transmission (and therefore less absorption) of x-rays through the patient. Even though the exposure per mAs increases as the kVp is increased (Fig. 23-9A), an accompanying reduction in the mAs will decrease the incident exposure to the patient. Unfortunately, there is a concomitant reduction in image contrast due to the higher effective energy of the x-ray beam. Within limits, this compromise is acceptable. Therefore, the patient exposure can be reduced by using a higher kVp and lower mAs.

Filtration of the polychromatic x-ray energy spectrum can significantly reduce exposure by selectively attenuating the low-energy x-rays in the beam that would otherwise be absorbed in the patient with little or no contribution to image formation. These low-energy x-rays mainly impart dose to the skin and shallow tissues where the beam enters the patient. As the tube filtration is increased, the beam becomes hardened (the effective energy increases) and the dose to the patient *decreases* because fewer low-energy photons are in the incident beam. The amount of filtration that can be added is limited by the increased tube loading necessary to offset the reduction in tube output, and the decreased contrast that occurs with excessive beam hardening. Except for mammography, a moderate amount of tube filtration does not significantly decrease subject contrast. The *quality* (effective energy) of the x-ray beam is assessed by measuring the HVL. State and federal regulations require the HVL of the beam to exceed a minimum value (see Chapter 5, Table 5-3). These HVL requirements apply to all diagnostic x-ray imaging equipment except mammography units. In the U.S., consideration is being given to increasing the required minimal filtration to match European standards. Ensuring adequate filtration is particularly relevant to avoiding or reducing skin injuries from prolonged fluoroscopically guided interventional procedures.

Field Area, Organ Shielding, and Geometry

Restriction of the field size by collimation to just the necessary volume of interest is an important dose reduction technique. In addition to limiting the patient volume exposed to the primary beam, it reduces the amount of scatter and thus the radiation dose to adjacent organs. From an image quality perspective, the scatter incident on the detector is also reduced, resulting in improved image contrast. When appropriate, shielding of critical organs should be used for patients undergoing a radiologic examination. For instance, when imaging a limb (such as a hand), a lap apron should be provided to the patient. In certain cases, *gonadal shielding* can be used to protect the gonads from primary radiation when the shadow of the shield does not interfere with the anatomy under investigation. Because the gonadal shield must attenuate primary radiation, its thickness must be equivalent to at least 0.5 mm of lead. In any situation, the use of patient protection must not interfere with the examination.

Increasing the source-to-object distance (SOD) and the source-to-image distance (SID) helps reduce dose. As the SOD and SID are increased, a reduced beam divergence limits the volume of the patient being irradiated, thereby reducing the integral dose. The exposure due to tube leakage is also reduced since the distance from the tube to the patient is increased, although this is only a minor consideration. For a fixed SID (e.g., C-arm fluoroscopic systems), patient dose is reduced by increasing the SOD as much as possible. Federal regulations specify a minimum patient (or tabletop) to focal spot distance of 20 cm. Furthermore, maintaining at least 30 cm is strongly recommended to prevent excessive radiation exposure.

X-Ray Image Receptors

The speed of the image receptor determines the number of x-ray photons and thus the patient dose necessary to achieve an appropriate signal level. Relative speed values are based on a par speed screen-film system, assigned a speed of 100; a higher speed system requires less exposure to produce the same optical density

(e.g., a 400-speed screen-film receptor requires four times less exposure than a 100-speed system). Specialized film processing can also affect the speed of the system. Either a faster screen or a faster film will reduce the incident exposure to the patient. In the case of a faster film, the disadvantage is increased quantum mottle and, in the case of a thicker intensifying screen, the disadvantage is reduced spatial resolution. Thus, increasing the speed of the screen-film image receptor reduces patient exposure, but the exposure reduction possible is limited by image quality concerns.

Computed radiography devices using photostimulable storage phosphor (PSP) detectors and other “digital” radiographic image receptors have wide exposure latitude. Subsequent image scaling (adjustments to brightness and contrast) and processing, allow the image to be manipulated for optimal viewing. Thus, these detectors can compensate (to some degree) for under- and overexposure and can reduce retakes caused by inappropriate radiographic techniques. However, underexposed images may contain excessive quantum mottle with poor contrast resolution, and overexposed images result in needless exposure to the patient. In comparison to screen-film receptors, PSP receptors are roughly equivalent to a 200-speed screen-film receptor in terms of quantum mottle and noise (e.g., for adult examinations of the abdomen and chest). Techniques for extremities with PSP receptors should be used at higher exposure levels (similar to extremity screen-film cassettes with speeds of 75 to 100), while exposures for pediatric examinations should be used at increased speed (e.g., ~400 speed) to reduce dose.

Conventional screen-film radiographic image receptors have an inherent safety feature. If the film processor is operating properly and the kVp is adequate, using an excessive quantity of radiation to produce an image results in an overly dark film. On the other hand, when digital image receptors are used, excessive incident radiation does not have a major adverse effect on image quality. Unnecessarily high exposures to patients may go unnoticed. Routine monitoring of the exposures to the digital image receptors is therefore necessary as a quality assurance concern.

The image intensifier is a detector system with a wide dynamic range, whereby the necessary entrance exposure is determined by a light-limiting aperture or the electronic gain of a subsequent detector (e.g., TV camera). Entrance exposures to the image intensifier as low as one μR per image for fluoroscopy or as high as 4 mR per image for digital subtraction angiography are possible. The amount of exposure reduction is limited by the quantum statistics (i.e., too little exposure will produce an image with excessive quantum mottle).

Fluoroscopy

Fluoroscopy imparts a very large portion of the dose delivered in diagnostic medical imaging examinations because of continuous x-ray production and real-time image output. Even though the exposure techniques are quite modest (e.g., 1 to 3 mA at 75 to 85 kVp for most studies), a fluoroscopic examination can require several minutes and, in some difficult cases, can exceed hours of “on” time. For example, a single-phase generator system at 2 mA at 80 kVp for 10 minutes of fluoro time (1,200 mAs) delivers an exposure of 6 R at 1 m (assuming 5 mR/mAs for 80 kVp at 1 m). If the focal spot is 30 cm from the skin, the patient entrance exposure is $(100/30)^2 \times 6 \text{ R} \approx 67 \text{ R}$! In the United States, fluoroscopic patient entrance exposure rates (30 cm from the x-ray source) are limited to a maximum of 10 R/min for systems with automatic brightness control (ABC) and to 5 R/min for systems with-

out ABC. Some systems have a high exposure rate fluoroscopic (“turbo”) mode providing up to 20 R/min output at 30 cm from the source, a factor of 10 times greater than the typical fluoro levels. Certainly, this capability must be used very judiciously. Using this mode requires continuous manual activation of a switch by the operator and a continuous audible signal indicates that this mode is being used. Cineangiography studies employ high exposure rates of 20 to 70 R/min, although with short acquisition times. The patient entrance exposure rate depends on the angulation of the beam through the patient, with the exposure rate increasing with the thickness of the body traversed by the beam [e.g., exposure rate greater for lateral projections than for anteroposterior (AP) or posteroanterior (PA) projections].

Serious x-ray–induced skin injuries to patients have been reported following prolonged fluoroscopically guided interventional procedures, such as percutaneous transluminal coronary angioplasty (PTCA), radiofrequency cardiac catheter ablation, vascular embolization, transjugular interhepatic portosystemic shunt placement, and percutaneous endovascular reconstruction. The cumulative skin doses in difficult cases can exceed 10 Gy (1,000 rads), which can lead to severe radiation-induced skin injuries (see Specific Organ System Responses—Skin in Chapter 25).

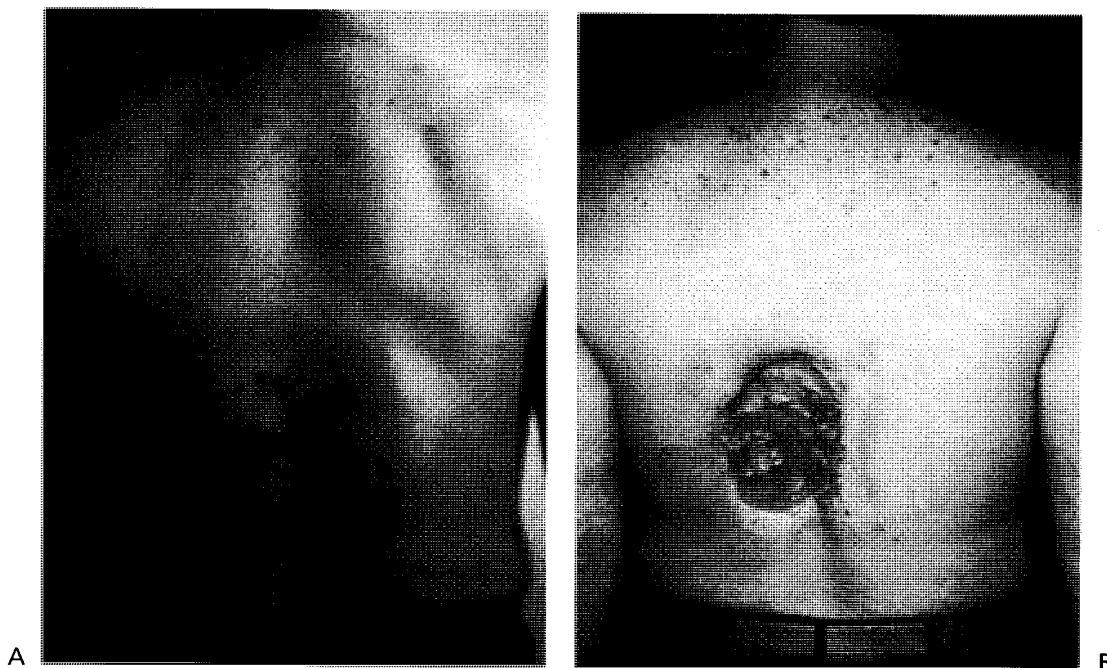


FIGURE 23-14. An example of a skin injury attributable to x-rays from fluoroscopy. This is a case of a 40-year-old man who underwent coronary angiography, coronary angioplasty, and a second angiography procedure due to complications, followed by a coronary artery bypass graft, all on March 29, 1990. **A:** The area of injury 6 to 8 weeks following the procedures. The injury was described as “turning red about 1 month after the procedure and peeling a week later.” In mid-May 1990, it had the appearance of a second-degree burn. In late summer 1990, exact date unknown, the appearance was that of a healed burn, except for a small ulcerated area present near the center. **B:** Skin breakdown continued over the following months with progressive necrosis. The injury eventually required a skin graft. While the magnitude of the skin dose received by this patient is not known, from the nature of the injury it is probable that the dose exceeded 20 Gy. (Adapted from Shope TB. Radiation-induced skin injuries from fluoroscopy. Rockville, MD: Center for Devices and Radiological Health, FDA. See <http://www.fda.gov/cdrh/rsnaii.html> for color images of this injury and additional information on radiation induced skin injuries from fluoroscopy.)

Figure 23-14 shows a severe skin injury following multiple fluoroscopic and angiographic procedures in which the total skin dose was estimated to be approximately 20 Gy (2,000 rads). Such injuries may not become fully apparent until weeks or even months after the procedure. The U.S. Food and Drug Administration (FDA) recommends that medical facilities determine the procedures likely to approach the threshold for skin injury and record, in the patient's medical record, the areas of the patient's skin that receive an absorbed dose that approaches or exceeds this threshold. The FDA also recommends training of physicians who operate or direct the operation of fluoroscopy systems, especially in the dose implications of the various modes of operation. While large radiation doses may be unavoidable in some of the more difficult cases, care should be taken during all examinations to employ techniques that will minimize the radiation exposure.

Reduction of fluoroscopic exposure is achieved in several ways. The most important method is to limit the "beam-on" time. The operator can significantly reduce fluoro time by using short bursts of exposure instead of staring at an unchanging static display. A 5-minute cumulative timer is required on all fluoro units to remind the operator audibly of each 5-minute time interval and to allow the technologist to keep track of the total amount of fluoro time for the examination. Last-image-hold devices using digital image memory are commonplace, and provide a continuous video display image after a short burst of x-rays. This type of device can reduce the fluoro time by 50% to 80% in many situations, particularly for physicians in training. *Pulsed fluoroscopy* reduces dose by allowing frame rates less than real time (e.g., 15 or 7.5 frames per second); it is used in conjunction with a digital image memory to provide a continuous video display with corresponding reductions in x-ray pulsing. This is an appropriate technology for situations not requiring high temporal resolution.

The use of an optimal image intensification system with appropriate conversion gain, high contrast ratio, and acceptable spatial resolution and contrast sensitivity also reduces patient dose. Restriction of the field size with collimation keeps the integral dose to the patient ALARA. Maintaining as much distance as possible between the x-ray tube and the patient's skin minimizes the skin dose. While magnification modes are helpful in improving spatial resolution, they increase the radiation dose significantly and should be used sparingly.

Devices that measure output exposure in fluoroscopic systems during procedures are called roentgen-area-product (RAP) or dose-area-product (DAP) meters. The detector of a RAP or DAP meter is a radiolucent ionization chamber that is positioned across the primary beam near the output port of the x-ray tube, but beyond the collimator. The meter measures the accumulated ionization that occurs during fluoroscopy and fluorography or cineradiography studies. Because the generated ionization signal is dependent on the collimated beam area, the exposure recorded by the device is not directly proportional to the entrance skin exposure to the patient. A calibration of the RAP meter response with changes in field of view, source-object distance, source-detector distance, and collimation are necessary to convert the accumulated exposure value after a procedure to the incident exposure received by the patient. An alternate and more accurate (yet more complex) method is to use a system that tracks kVp, mA, SOD, SID, and table position continuously during the examination. Exposure outputs as a function of all output variables can be measured initially on phantoms, and used to calibrate the exposure monitoring system. In any exposure monitoring system, the real-time display of accumulated patient exposure provides

the physician with the ability to modify the procedure, if necessary, to minimize the potential for radiation-induced skin injuries. The total exposure recorded at the completion of the procedure also allows the physician to identify patients who may be at risk for radiation-induced skin damage and thus should be followed.

Physicians using fluoroscopic systems should be trained and credentialed to assure they have an understanding of the techniques used to minimize patient and personnel exposure, particularly the dose implications of the different modes of operation. In addition, physicians should be able to demonstrate knowledge of current radiation control regulations.

Computed Tomography

Most radiographic images, other than those acquired with portable machines, are acquired utilizing automatic exposure control (phototiming). In CT, however, automatic exposure modes are usually not available. Consequently, for a given CT protocol (e.g., abdomen), most CT scanners use a fixed kVp and mAs regardless of the size of the patient, unless the CT technologist changes it. Some scanners now have an ability to modulate the mA as a function of tube rotation angle, to compensate for the thickness variation that occurs in the torso of the body from the AP to the lateral projections, which can markedly reduce exposure without decreasing image quality. The fixed techniques in CT for a given protocol often lead to unnecessarily high mAs values for smaller patients, particularly neonatal and young children. The only way to avoid this unnecessary radiation dose to the patient is to train the CT technologists to reduce the mAs (and perhaps the kVp) for thinner patients. A technique chart should be devised for each CT scanner showing kVp and mAs values for different diameter patients as a guide to achieving good image quality at proper dose levels. Table 23-16 shows the dose reductions that can be achieved if the CT

TABLE 23-16. DOSE REDUCTION POTENTIAL IN COMPUTED TOMOGRAPHY (CT)

Patient diameter (cm)	Percent mAs	mAs
14	0.3	1
16	0.6	2
18	1.2	4
20	2.2	7
22	4.2	13
24	7.9	24
26	15	45
28	28	85
30	53	160
32	100	300
34	188	564
36	352	1,058
38	661	1,983
40	1237	3,712

Note: If a CT technique of 120 kVp and 300 mAs delivers good image quality (low noise images) for a 32-cm-diameter patient, the same image quality (SNR) can be achieved by reducing the mAs to the levels indicated for patients smaller than 32, and increasing the mAs for patients larger than 32 cm.

protocols (mAs values) are adjusted to the patient's size. The table was developed assuming that a technique of 120 kVp and 300 mAs (5 mm slice thickness and pitch = 1.0) delivers good image quality for an average 32 cm diameter torso. (These are approximate values and techniques for different scanners and institutions will vary). Table 23-16 was calculated so that the number of x-ray photons reaching the central CT detector remains constant (i.e., constant SNR). For a 30-cm patient, approximately half of the dose (160 mAs instead of 300 mAs) can be used while maintaining the same image quality. These large reductions in dose are achievable because x-ray attenuation is an *exponential* phenomenon. For a pediatric patient with a small 20 cm diameter torso, only 7 mAs is needed to produce the same image quality as 300 mAs for a 32-cm patient. This represents a factor of *43 times less dose to the pediatric patient* for the same signal-to-noise ratio (SNR)! Even if twice the SNR is needed to visualize anatomic changes in the pediatric patient, 20 times less dose can still be achieved, an extremely important consideration for the radiation-sensitive young child.

Miscellaneous Considerations

Careful identification of the patient and, for female patients of reproductive age, determination of pregnancy status are necessary before an examination is performed. A significant reduction in population dose can be achieved by eliminating screening exams that only rarely detect pathology. Standing-order x-rays, such as presurgical chest exams for hospital admissions or surgeries not involving the chest, are inappropriate. Frequent screening examinations are often not indicated without a specific reason or a reasonable amount of time elapsed (e.g., years) from the previous exam. A good example is the "yearly" dental x-ray. While yearly exams are appropriate for some patients, a 2-year interval between x-ray examinations in non-cavity-prone adult patients with a healthy mouth is appropriate. Another opportunity for significant patient dose reduction in dentistry is the use of high-speed film (e.g., E-speed film) or digital image receptors. Many dental offices still use slower D-speed films that require approximately twice the exposure (and thus patient dose). Periodic screening mammography examinations are not appropriate for women younger than 35 to 40 years old, unless there is a familial history of breast cancer or other indication. However, as the probability of cancer increases with age, and the fact that survival is drastically enhanced by early detection makes yearly screening mammograms sensible for the general population of women above 40 to 50 years old.

The frequency of repeat exams due to faulty technique or improper position is primarily determined by the skill and diligence of the x-ray technologists. The average repeat rate is approximately 4% to 6%, with a range from 1% to 15%. Training institutions often have repeat rates toward the higher end of the range, mainly due to lack of experience. The largest numbers of films that must be repeated are from portable exams, in which positioning difficulty causes the anatomy of interest to be improperly represented on the film, and lack of automatic exposure control increases the likelihood of improper optical densities. Preprogrammed radiographic techniques for examinations are often available on radiographic equipment, and can eliminate the guesswork in many radiographic situations (of course, provided that the techniques are properly set up in the first place). Technique charts for various examinations should be posted conspicuously at the control panel to aid the tech-

nologist in the correct selection of radiographic technique. The use of photostimulable phosphor imaging plates (Chapter 11) can significantly reduce the number of exams that must be repeated because of improper radiographic technique. The examinations with the highest retake rates are portable chests; lumbar spines; thoracic spines; kidneys, ureters, and bladder (KUBs); and abdomens. Continuous monitoring of retakes and inadequate images and identification of their causes are important components of a quality assurance program, so that appropriate action may be taken to improve quality.

Equipment problems, such as improperly loaded cassettes, excessive fog due to light leaks or poor film storage conditions, processor artifacts caused by dirty components or contaminated chemicals, uncalibrated x-ray equipment, or improper imaging techniques can also lead to retakes. Many of these errors are eliminated by a periodic quality control program that tests the performance of the x-ray equipment, image receptors, and processing systems.

Radiation Safety in Nuclear Medicine

Contamination Control and Surveys

Contamination is simply uncontained radioactive material located where it is not wanted. Contamination control methods are designed to prevent radioactive material from coming into contact with personnel and to prevent its spread to other work surfaces. Protective clothing and handling precautions to control contamination are similar to the “universal precautions” that are taken to protect hospital personnel from pathogens. In most cases, disposable plastic gloves, laboratory coats, and closed-toed shoes offer adequate protection. One advantage of working with radioactive material, in comparison to other hazardous substances, is that small quantities can be easily detected. Personnel and work surfaces must be routinely surveyed for contamination. Nonradioactive work areas near radioactive material use areas (e.g., reading rooms and patient exam areas) should be posted as “clean areas” where radioactive materials are prohibited. Work surfaces where unsealed radioactive materials are used should be covered by plastic-backed absorbent paper that is changed when contaminated or worn. Volatile radionuclides (e.g., I-131 and Xe-133 gas) should be stored in a 100% exhaust fume hood to prevent airborne contamination and subsequent inhalation. The collimators of scintillation cameras should be covered with plastic to avoid contamination that could render the cameras unusable until the radionuclide has decayed. Personnel should discard gloves into designated radioactive waste receptacles after working with radioactive material and monitor their hands, shoes, and clothing for contamination at periodic intervals. All personnel should wash their hands after handling radioactive materials and especially before eating or drinking to minimize the potential for internal contamination. If the skin becomes contaminated, the best method of decontamination is washing with soap and warm water. The skin should not be decontaminated too aggressively to avoid creating abrasions that can enhance internal absorption. External contamination is rarely a serious health hazard; however, internal contamination can lead to significant radiation exposures. Good contamination control techniques help prevent inadvertent internal contamination.

The effectiveness of contamination control is monitored by periodic Geiger-Mueller (GM) meter surveys (whenever contamination is suspected and at the end

of each workday) and wipe tests (typically weekly) of radionuclide use areas. Wipe tests are performed by using small pieces of filter paper, alcohol wipes, or cotton-tipped swabs to check for removable contamination at various locations throughout the nuclear medicine laboratory. These wipe test samples (called "swipes") are counted in a NaI (Tl) gamma well counter. Areas that are demonstrated to be in excess of twice background radiation levels are typically considered to be contaminated. The contaminated areas are immediately decontaminated followed by additional swipes to confirm the effectiveness of decontamination efforts. GM meter surveys are also performed throughout the department to detect areas of contamination. In addition, exposure rate measurements (with a portable ion chamber) are made near areas that could produce high exposure rates (e.g., radioactive waste and material storage areas).

Laboratory Safety Practices

The following are other laboratory safety practices that minimize the potential for personnel or facility contamination:

1. Label all radioactive materials containers with the radionuclide, measurement or calibration date, activity, and chemical form.
2. Do not eat, drink, or smoke in the radioactive work areas.
3. Do not pipette radioactive materials by mouth.
4. Discard all radioactive materials into radioactive waste disposal containers.
5. Ensure that Xe-133 ventilation studies are performed in a room with negative pressure with respect to the hallway. (Negative pressure will occur if the room air exhaust rate substantially exceeds the supply rate. The difference is made up by air flowing from the hallway into the room, preventing the escape of the xenon gas.) Xenon exhaled by patients undergoing ventilation studies must be exhausted into a Xe trap. Xe traps are large shielded activated charcoal cartridges, which adsorb xenon on the large surface areas presented by the charcoal, thus slowing its migration through the cartridge and allowing significant decay to occur before it is exhausted to the environment.
6. Report serious spills or accidents to the radiation safety officer or health physics staff.

Radioactive Material Spills

First aid takes priority over decontamination after an accident, and personnel decontamination takes priority over facility decontamination efforts. Individuals in the immediate area should be alerted to a spill and remain until they can be monitored with a GM survey instrument to assure that they have not been contaminated. Radioactive material spills should be contained with absorbent material and the area isolated and posted with warning signs indicating the presence of radioactive contamination. Disposable shoe covers and plastic gloves should be donned before beginning decontamination. Decontamination should be performed from the perimeter of the spill toward the center to limit the spread of contamination. Decontamination is usually accomplished by simply absorbing the spill and cleaning the affected areas with detergent and water. A wipe test and meter survey should be used to verify successful decontamination. Personnel participating in decontam-

ination should remove their protective clothing and be surveyed with a GM survey instrument to assure that they are not contaminated. If the spill involves volatile radionuclides or if other conditions exist that suggest the potential for internal contamination, bioassays should be performed by the radiation safety staff. Common bioassays for internal contamination include external counting with a NaI (Tl) detector (e.g., thyroid bioassay for radioiodine) and radioactivity measurement of urine (e.g., tritium bioassay).

Protection of the Patient in Nuclear Medicine

Sometimes nuclear medicine patients are administered the wrong radiopharmaceutical or the wrong activity. These accidents are called medical events or misadministrations. The following event has occurred at a several institutions: a patient referred for a whole-body bone scan [740 MBq (20 mCi) of Tc-99m MDP] has been mistakenly administered up to 370 MBq (10 mCi) of I-131 sodium iodide for a whole-body thyroid cancer survey [thyroidal dose for 370 MBq and 20% uptake is ~100 Gy (10,000 rad)]. The cause has often been a verbal miscommunication. In another case, a mother who was nursing an infant was administered a therapeutic dose of I-131 sodium iodide without being instructed to discontinue breast-feeding, resulting in an estimated dose of 300 Gy (30,000 rad) to the infant's thyroid. The following precautions, most of which are mandated by the U.S. Nuclear Regulatory Commission (NRC), are intended to reduce the frequency of such incidents.

The vial radiation shield containing the radiopharmaceutical vial must be conspicuously labeled with the name of the radiopharmaceutical or its abbreviation. Each syringe containing a dosage of a radiopharmaceutical must be conspicuously labeled with the patient's name and the radiopharmaceutical (or its abbreviation). The patient's identity must be verified, whenever possible by two means (e.g., by having the patient recite his or her name and social security number). The possibility that a woman is pregnant or breast-feeding must be ascertained before administering the radiopharmaceutical. In the case of activities of I-131 sodium iodide exceeding 30 μ Ci and therapeutic radiopharmaceuticals, pregnancy should be ruled out by a pregnancy test, certain premenarche in a child, certain postmenopausal state, or documented hysterectomy or tubal ligation.

Before the administration of activities exceeding 1.1 MBq (30 μ Ci) of I-131 in the form of sodium iodide and any radionuclide therapy, a written directive must be prepared by the nuclear medicine physician identifying the patient, the radionuclide, the radiopharmaceutical, the activity to be administered, and the route of administration. The written directive must be consulted at the time of administration and the patient's identity should be verified by two methods. Although not required by the NRC, similar precautions should be taken when reinjecting autologous radiolabeled blood products to prevent transfusion reactions and the transmission of pathogens.

Women who are nursing infants by breast at the time of the examination should be counseled to discontinue breast-feeding until the radioactivity in the breast milk has been reduced to a safe level. Table 23-17 contains recommendations for cessation of breast-feeding after the administration of radiopharmaceuticals to mothers. The NRC requires written instructions to nursing mothers if total effective dose equivalent to a nursing infant could exceed 1 mSv (100 mrem) (Code of

TABLE 23-17. RECOMMENDATIONS FOR CESSATION OF BREAST FEEDING AFTER ADMINISTRATION OF RADIOPHARMACEUTICALS TO MOTHERS

Radiopharmaceuticals	Administered activity ^a	Imaging procedure	Safe breast milk concentration (μCi/ml)	Cessation of breast feeding until breast milk is safe
Tc-99m sodium pertechnetate	10 mCi	Thyroid scan and Meckel's scan	8.2×10^{-2}	24 hr
Tc-99m kits (general rule)	5–25 mCi	All	8.2×10^{-2}	24 hr
Tc-99m DTPA	10–15 mCi	Renal scan	1.2×10^{-1}	17 hr
Tc-99m MAA	3–5 mCi	Lung perfusion scan	1.2×10^{-1}	10 hr
Tc-99m sulfur colloid	5 mCi	Liver spleen scan	1.6×10^{-1}	15 hr
Tc-99m MDP	15–25 mCi	Bone scan	2.1×10^{-1}	17 hr
Ga-67 Citrate	6–10 mCi	Infection and tumor scans	2.1×10^{-3}	4 wk
Tl-201 Chloride	3 mCi	Myocardial perfusion	2.4×10^{-3}	3 wk
Sodium I-123	30 μCi	Thyroid uptake	1.2×10^{-4}	3 days
	50–400 μCi	Thyroid scan	"	5 days
Sodium I-131	5 μCi	Thyroid uptake	4.1×10^{-7}	68 days
"	10 mCi	Thyroid cancer Scan or Graves therapy	"	Discontinue ^b
"	33 mCi	Outpatient therapy for hyperfunctioning nodule	"	Discontinue ^b
"	100 mCi or more	Thyroid cancer treatment (ablation)	"	Discontinue ^b

^aMultiply by 37 to obtain MBq.

^bDiscontinuance is based not only on the excessive time recommended for cessation of breast feeding but also on the high dose the breasts themselves would receive during the radiopharmaceutical breast transit.

DTPA, diethylenetriamine pentaacetic acid; MAA, macroaggregated albumin; MDP, methylene diphosphate.

Adapted from Conte AC, Bushberg JT. Essential science of nuclear medicine. In: Brant WE, Helms CA, eds. *Fundamentals of diagnostic radiology* 2nd. ed. Baltimore: Williams & Wilkins, 1999.

Federal Regulation Title 10 Part 35.75.) The instructions include guidance on the interruption or discontinuation of breast-feeding, and information on the potential consequences, if any, of failure to follow the guidance.

Radionuclide Therapy

Treatment of thyroid cancer and hyperthyroidism with I-131 sodium iodide is a proven and widely utilized form of radionuclide therapy. NaI-131 is manufactured in liquid and capsule form. I-131 in capsules is incorporated into a waxy or crystalline matrix or bound to a gelatin substance, all of which reduce the volatility of the I-131. In recent years, advances in monoclonal antibody and chelation chemistry technology have produced a variety of radioimmunotherapeutic agents, many

labeled with I-131. I-131 decays with an 8-day half-life and emits high-energy beta particles and gamma rays. These decay properties, together with the facts that I-131 can be released as a gas under certain conditions, can be absorbed through the skin, and concentrates and is retained in the thyroid, necessitate a number of radiation protection precautions with the administration of therapeutic quantities of I-131.

Following administration, I-131 is secreted/excreted in all body fluids including urine, saliva, and perspiration. Before I-131 is administered, surfaces of the patient's room likely to become contaminated, such as the floor, bed controls, mattress, light switches, toilet, and telephone, are covered with plastic or absorbent plastic-backed paper to prevent contamination. In addition, the patient's meals are served on disposable trays. Waste containers are placed in the patient's room to dispose of used meal trays and to hold contaminated linens for decay. Radiation safety staff will measure exposure rates at the bedside, at the doorway, and in neighboring rooms. In some cases, immediately adjacent rooms must be posted off limits to control radiation exposure to other patients and nursing staff. These measurements are posted, together with instructions to the nurses and visitors including maximal permissible visiting times. Nursing staff members are required to wear dosimeters and are trained in the radiation safety precautions necessary to care for these patients safely and efficiently. Visitors are required to wear disposable shoe covers, and staff members wear both shoe covers and disposable gloves to prevent contamination. Visiting times are limited and patients are instructed to stay in their beds and avoid direct physical contact to keep radiation exposure and contamination to a minimum. Visitors and staff are usually restricted to nonpregnant adults. After the patient is discharged, the health physics or nuclear medicine staff decontaminates the room and verifies through wipe tests and GM surveys that the room is completely decontaminated. Federal or state regulations may require thyroid bioassays of staff technologists and physicians directly involved with dose preparation or administration of large activities of I-131.

The NRC regulations (Title 10 Part 35.75) require that patients receiving therapeutic radionuclides be hospitalized until or unless it can be demonstrated that the total effective dose equivalent to any other individual from exposure to the released patient is not likely to exceed 5 mSv (0.5 rem). Guidance on making this determination can be found in two documents from the NRC: NUREG-1556, Vol. 9, entitled "A Consolidated Guidance about Materials Licenses: Program-Specific Guidance about Medical Licenses" and Regulatory Guide 8.39, entitled "Release of Patients Administered Radioactive Materials." These documents describe methods for calculating doses to other individuals and contain tables of activities not likely to cause doses exceeding 5 mSv (0.5 rem). For example, patients may be released from the hospital following I-131 therapy when the activity in the patient is at or below 1.2 GBq (33 mCi) or when the dose rate at 1 m from the patient is at or below 0.07 mSv/hr (7mrem/hr).

To monitor the activity of I-131 in a patient, an initial exposure measurement is obtained at 1 m from the patient. This exposure rate is proportional to the administered activity. Exposure rate measurements are repeated daily until the exposure rate associated with 1.2 GBq (33 mCi) is obtained. The exposure rate from 33 mCi in a patient will vary with the mass of the patient. The activity remaining in the patient can be estimated by comparing initial and subsequent exposure rate measurements made at a fixed, reproducible geometry. The exposure rate equivalent to 1.2 GBq (33 mCi) remaining the patient ($X_{1.2}$) is:

$$X_{1.2} = \frac{(1.2 \text{ GBq})E_0}{A_0} \quad [23-12]$$

where

E_0 = initial exposure rate (~15 minutes after radiopharmaceutical administration)
and

A_0 = administered activity (GBq).

For example, a thin patient is administered 5.55 GBq (150 mCi) I-131 after which an exposure rate measurement at 1 m reads 37 mR/hr. The patient can be discharged when the exposure rate falls below:

$$X_{1.2} = \frac{(1.2 \text{ GBq})(37 \text{ mR/hr})}{5.55 \text{ GBq}} = 8.0 \text{ mR/hr}$$

Once below 1.2 GBq (33 mCi), assuming a stable medical condition, the patient may be released. If the total effective dose equivalent to any other individual is likely to exceed 1 mSv (0.1 rem), the NRC requires the licensee to provide the released individual, or the individual's parent or guardian, with radiation safety instructions. These instructions must be provided in writing and include recommended actions that would minimize radiation exposure to, and contamination of, other individuals. Typical radiation safety precautions for home care following radioiodine therapy are shown below. These precautions also apply to patient therapy with less than 1.2 GBq (33 mCi) of I-131 for hyperthyroidism. On the rare occasions when these patients are hospitalized for medical reasons, contamination control procedures similar to thyroid cancer therapy with I-131 should be observed.

The following is an example of typical radiation safety instructions that are given to the patient.

General Safety Guide For Home Care Following Radioiodine Therapy: 370 MBq–1.2 GBq (10–33 mCi)

The dose of radioiodine that you have received is beneficial to you but it is desirable that other persons not be unnecessarily exposed to radiation. Below we have suggested some points, which will help keep such exposure to others as low as possible.

Radiation Safety Precautions Are Necessary

During the First Four Days

The majority of the remaining activity in your body will be eliminated through the urine. In addition, a small portion of radioactivity will be found in your saliva and perspiration. To minimize the spread of radioactivity:

1. Wash cups, plates, and eating utensils immediately after use, or use disposables.
2. Do not kiss anybody.
3. Use individual towels and washcloths.
4. Sleep in a separate bed during this period.
5. At the end of two days, all personal clothing with which you have come in contact (pajamas, underwear, towels, bed linens, etc.) should be washed separately from those used by other members of the family.
6. Contact with infants and pregnant women should be avoided during this four-day period.

7. Stay three feet away or further from other people, except for very brief contact.

A large portion of the radioactivity is excreted in the urine during the first two days after administration. Double-flush after using the toilet. If urine is spilled, splashed on toilet seat, etc., wash and rinse the affected area three times, using disposable paper or tissue, and flush. You may prepare food after washing your hands.

During the Remainder of the Week

Your thyroid gland will contain significant levels of radioactivity. To minimize external exposure, avoid sitting close (within a foot) to others for hours at a time (e.g., in a theater). You need not be concerned about being close to people for short periods of time (a few minutes). Avoid holding infants or young children for long periods each day.

Breast-Feeding

Breast-feeding must be discontinued. Your physician will advise you when you may resume.

Phosphorus-32 is used for the radionuclide therapy of polycythemia vera, malignant effusion, and other diseases. Strontium-89 chloride is used to treat intractable pain from metastatic bone disease. The primary radiation safety precaution for these radionuclide therapies is contamination control. For example, the wound and bandages at the site of an intraabdominal instillation of P-32 should be checked regularly for leakage. The blood and urine will be contaminated and universal precautions should be observed. Exposure rates from patients treated with pure beta emitters like P-32 are not significant. Likewise the exposure rate from patients treated with Sr-89 are insignificant compared to the risk of contamination.

Radiation Safety Program

Radioactive material use regulations require “cradle to grave” control of all radiation sources. A license from a state or federal regulatory agency is required to possess and use radioactive materials in medicine. The license will specify a number of program elements, including the operational details of the radiation safety program. The license will specify the individual responsible for assuring compliance with all radiation safety regulations (called the “radiation safety officer”), as well as procedures and record-keeping requirements for radioactive material receipt, transportation, use, and disposal. Larger institutions will have a health physics department and a Radiation Safety Committee that oversees the radiation safety program. The Radiation Safety Committee comprises the radiation safety officer and representatives from departments with substantial radiation and radioactive material use (e.g., nuclear medicine, diagnostic radiology, radiation oncology, and nursing) as well as representation from hospital administration.

Radioactive Waste Disposal

Minimizing radioactive waste is an increasingly important element of radioactive material use. Fortunately, most radionuclides used in nuclear medicine have short half-lives that allow them to be held until they have decayed. As a rule, radioactive material is held for at least 10 half-lives and then surveyed with an appropriate radi-

ation detector (typically a GM survey instrument with a pancake probe) to confirm the absence of any significant radioactivity before being discarded as nonradioactive waste. Disposal of decayed radioactive waste in the nuclear medicine department is made easier by segregating the waste into short half-life (e.g., F-18 and Tc-99m), intermediate half-life (e.g., Ga-67, In-111, I-123, Xe-133, Tl-201), and long half-life (e.g., P-32, Cr-51, Sr-89, I-131) radionuclides.

A small amount of radioactive material may be disposed of in the sink as long as the material is water soluble and the total amount does not exceed regulatory limits. Detailed records of all radioactive material disposals must be maintained for inspection by regulatory agencies. Radioactive excreta from patients receiving diagnostic or therapeutic radiopharmaceuticals are exempt from these disposal regulations and thus may be disposed into the sanitary sewer. Thus, I-131 used for thyroid cancer therapy and all diagnostic radiopharmaceuticals that are excreted by the patient are exempt from environmental release regulations. The short half-lives and large dilution in the sewer system provide a large margin of safety with regard to environmental contamination and public exposure.

23.5 REGULATORY AGENCIES AND RADIATION EXPOSURE LIMITS

Regulatory Agencies

A number of regulatory agencies have jurisdiction over various aspects of the use of radiation in medicine. The regulations promulgated under their authority carry the force of law. These agencies can inspect facilities and records, levy fines, suspend activities, and issue and revoke radiation use authorizations.

The U.S. Nuclear Regulatory Commission (NRC) regulates *special nuclear* material (plutonium and uranium enriched in the isotopes U-233 and U-235), *source* material (thorium, uranium, and their ores) and *by-product* material of nuclear fission (e.g., fission products) used in the commercial nuclear power industry, research, medicine, and a variety of other commercial activities. Under the terms of the Atomic Energy Act of 1954, which established the Atomic Energy Commission whose regulatory arm now exists as the NRC, the agency was given regulatory authority only for special nuclear, source, and by-product material. Thus, radioactive materials produced by means other than nuclear fission (e.g., cyclotron-produced radionuclides such as F-18, Tl-201, I-123, and In-111) are not subject to NRC regulation. Some states administer their own radiation control programs for radioactive materials and other radiation sources. These states have an agreement with the NRC to promulgate and enforce regulations similar to those of the NRC. In addition, these programs typically regulate all sources of radiation including cyclotron-produced radionuclides and radiation-producing machines. These "agreement states" carry out the same regulatory supervision, inspection, and enforcement actions that would otherwise be performed by the NRC.

While NRC and agreement state regulations are not identical, there are some essential aspects that are common to all of these regulatory programs. Workers must be informed of their rights and responsibilities, including the risks inherent in utilizing radiation sources and their responsibility to follow established safety procedures. The NRC's regulations are contained within Title 10 (Energy) of the Code of Federal Regulations (CFR). The most important sections for medical use of radionu-

clides are the “Standards for Protection Against Radiation” (Part 20) and “Medical Use of By-Product Material” (Part 35). Part 20 contains the definitions utilized in the radiation control regulations and requirements for radiation surveys; personnel monitoring (dosimetry and bioassay); radiation warning signs and symbols; and shipment, receipt, control, storage, and disposal of radioactive material. Part 20 also specifies the maximal permissible doses to radiation workers and the public; environmental release limits; and documentation and notification requirements after a significant radiation accident or event, such as the loss of a brachytherapy source or the release of a large quantity of radioactive material to the environment.

Part 35 lists the requirements for the medical use of by-product material. Some of the issues covered in this section include the medical use categories, training requirements, precautions to be followed in the medical use of radiopharmaceuticals, testing and use of dose calibrators, and requirements for the reporting of medical events (misadministrations) of radiopharmaceuticals. The NRC also issues regulatory guidance documents, which provide the licensee with methods acceptable to NRC for satisfying the regulations. The procedures listed in these documents may be adopted completely or in part with the licensee’s own procedures, which will be subject to NRC review and approval.

The FDA regulates radiopharmaceutical development and manufacturing as well as the performance and radiation safety requirements associated with the production of commercial x-ray equipment. Although this agency does not directly regulate the end user (except for mammography), it does maintain a strong involvement in both the technical and regulatory aspects of human research with radioactive materials and other radiation sources as well as publish documents in areas of interest including radiologic health, design and use of x-ray machines, and radiopharmaceutical development.

The U.S. Department of Transportation (DOT) regulates the transportation of radioactive materials. Other regulations and recommendations related to medical radiation use programs are promulgated by other federal, state, and local agencies.

Advisory Bodies

Several advisory organizations exist that periodically review the scientific literature and issue recommendations regarding various aspects of radiation protection. While their recommendations do not constitute regulations and thus do not carry the force of law, they are usually the origin of the most of the regulations adopted by regulatory agencies and are widely recognized as “standards of good practice.” Many of these recommendations are voluntarily adopted by the medical community even in the absence of a specific legal requirement. The two most widely recognized advisory bodies are the National Council on Radiation Protection and Measurements (NCRP) and the International Commission on Radiological Protection (ICRP). The NCRP is a nonprofit corporation chartered by Congress to collect, analyze, develop, and disseminate, in the public interest, information and recommendations about radiation protection, radiation measurements, quantities, and units. In addition, it is charged with working to stimulate cooperation and effective utilization of resources regarding radiation protection with other organizations including the ICRP.

The ICRP is similar in scope to the NCRP; however, its international membership brings to bear a variety of perspectives on radiation health issues. The NCRP and ICRP have published over 200 monographs containing recommenda-

tions on a wide variety of radiation health issues that serve as the reference documents from which many regulations are crafted.

NRC “Standards for Protection Against Radiation”

As mentioned above, the NRC has established “Standards for Protection Against Radiation” (10 CFR 20) to protect radiation workers and the public. These regulations were extensively revised in 1991, at which time many recommendations of the ICRP, the NCRP, and the regulated community were adopted. The revised regulations incorporate a twofold system of dose limitation: (a) the doses to individuals shall not exceed limits established by the NRC, and (b) all exposures shall be kept as low as reasonably achievable (ALARA), social and economic factors being taken into account. The new regulations also adopt a system recommended by the ICRP, which permits internal doses (from ingested or inhaled radionuclides) and external doses to be summed, with a set of limits for the sum.

Summing Internal and External Doses

There are significant differences between external and internal exposures. The dose from an internal exposure continues after the period of ingestion or inhalation, until the radioactivity is eliminated by radioactive decay or biologic removal. The exposure may last only a few minutes, for example, in the case of the radionuclide O-15 ($T_{1/2} = 122$ seconds), or may last the lifetime of the individual, as is the case for the ingestion of long-lived Ra-226. The *committed dose equivalent* ($H_{50,T}$) is the dose equivalent to a tissue or organ over the 50 years following the ingestion or inhalation of radioactivity. The committed effective dose equivalent (CEDE) is a weighted average of the committed dose equivalents to the various tissues and organs of the body:

$$\text{CEDE} = \sum w_T H_{50,T}$$

The NRC adopted the tissue weighting factors (w_T) from ICRP reports 26 and 30, which predate those in the most current ICRP recommendations (ICRP report 60, see Chapter 3: Effective Dose).

To sum the internal and external doses to any individual tissue or organ, the deep dose equivalent (indicated by a dosimeter worn by the exposed individual) and the committed dose equivalent to the organ are added. The sum of the external and internal doses to the entire body, called the total effective dose equivalent (TEDE), is the sum of the deep dose equivalent and the committed effective dose equivalent.

Dose Limits

The NRC’s radiation dose limits are intended to limit the risks of stochastic effects, such as cancer and genetic effects, and to prevent deterministic effects, such as cataracts, skin damage, sterility, and hematologic consequences of bone marrow depletion (deterministic and stochastic effects are discussed in Chapter 25). To limit the risk of stochastic effects, the sum of the external and internal doses to the entire body, the total effective dose equivalent, may not exceed 50 mSv/yr (5 rem/yr). To prevent deterministic effects, the sum of the external dose and committed dose equivalent to any individual organ (except the lens of the eye) may not exceed 500

TABLE 23-18. NUCLEAR REGULATORY COMMISSION (NRC) REGULATORY REQUIREMENTS: MAXIMUM PERMISSIBLE DOSE EQUIVALENT LIMITS^a

Limits	Maximum permissible annual dose limits	
	mSv	rem
Occupational limits		
Total effective dose equivalent	50	5
Total dose equivalent to any individual organ (except lens of eye)	500	50
Dose equivalent to the lens of the eye	150	15
Dose equivalent to the skin or any extremity	500	50
Minor (<18 years old)	10% of adult limits	10% of adult limits
Dose to an embryo/fetus ^b	5 in 9 months	0.5 in 9 months
Nonoccupational (public limits)		
Individual members of the public	1.0/yr	0.1/yr
Unrestricted area	0.02 in any 1 hr ^c	0.002 in any 1 hr ^c

^aThese limits are exclusive of natural background and any dose the individual has received for medical purposes; inclusive of internal committed dose equivalent and external effective dose equivalent (i.e., total effective dose equivalent).

^bApplies only to conceptus of a worker who declares her pregnancy. If the limit exceeds 4.5 mSv (450 mrem) at declaration, conceptus dose for remainder of gestation is not to exceed 0.5 mSv (50 mrem).

^cThis means the dose to an area (irrespective of occupancy) shall not exceed 0.02 mSv (2 mrem) in any 1 hour. This is not a restriction of instantaneous dose rate to 0.02 mSv/hr (2 mrem/hr).

mSv/yr (50 rem/yr). The dose to the fetus of a declared pregnant radiation worker may not exceed 5 mSv (0.5 rem) over the 9-month gestational period and should not substantially exceed 500 μ Sv (50 mrem) in any one month. Table 23-18 lists the most important NRC radiation exposure limits.

Annual Limits on Intake and Derived Air Concentrations

In practice, the committed dose equivalent and committed effective dose equivalent are seldom used in protecting workers from ingesting or inhaling radioactivity. Instead, the NRC has established annual limits on intake (ALIs), which limit the inhalation or ingestion of radioactive material to activities that will not cause radiation workers to exceed any of the limits in Table 23-18. The ALIs are calculated for the “standard person” and expressed in units of microcuries. ALIs are listed for the inhalation and oral ingestion pathways for a wide variety of radionuclides (Table 23-19).

TABLE 23-19. EXAMPLES OF DERIVED AIR CONCENTRATIONS (DAC) AND ANNUAL LIMITS ON INTAKE (ALI) FOR OCCUPATIONAL EXPOSED WORKERS (1991 NRC REGULATIONS: 10 CFR 20)

Radionuclide	DAC (μ Ci/mL) ^a	ALI (μ Ci) ^a Ingestion
I-125	3×10^{-8}	40
I-131	2×10^{-8}	30
Tc-99m	6×10^{-5}	8×10^4
Xe-133 (gas)	1×10^{-4}	N/A

^aMultiply μ Ci/mL (or μ Ci) by 37 to obtain kBq/mL (or kBq).
N/A, not applicable.

Exposure to airborne activity is also regulated via the derived air concentration (DAC), which is that concentration of a radionuclide that, if breathed under conditions of light activity for 2,000 hours (the average number of hours worked in a year), will result in the ALI; DACs are expressed in units of $\mu\text{Ci/mL}$. Table 23-19 lists DACs and ALIs for volatile radionuclides of interest in medical imaging. In circumstances in which either the external exposure or the internal deposition does not exceed 10% of its respective limits, summation of the internal and external doses is not required to demonstrate compliance with the dose limits; most diagnostic radiology and nuclear medicine departments will be able to take advantage of this exemption.

As Low As Reasonably Achievable (ALARA) Principle

Dose limits to workers and the public are regarded as upper limits rather than as acceptable doses or thresholds of safety. In fact, the majority of occupational exposures in medicine and industry result in doses far below these limits. In addition to the dose limits specified in the regulations, all licensees are required to employ good health physics practices and implement radiation safety programs to ensure that radiation exposures are kept *as low as reasonably achievable* (ALARA), taking societal and economic factors into consideration. This ALARA doctrine is the driving force for many of the policies, procedures, and practices in radiation laboratories, and represents a commitment by both employee and employer to minimize radiation exposure to staff, the public, and the environment to the greatest extent practical. For example, the *L*-shield in the nuclear pharmacy may not be specifically required by a regulatory authority; however, its use has become common practice in nuclear medicine departments and represents a “community standard” against which the departments’ ALARA programs will be evaluated. The application of the ALARA principle to diagnostic x-ray room shielding was presented above (see Shielding in Diagnostic Radiology).

SUGGESTED READING

- American Association of Physicists in Medicine. *Managing the use of fluoroscopy in medical institutions*. AAPM Radiation Protection Task Group no. 6. AAPM report no. 58. Madison, WI: Medical Physics, 1998.
- Archer BR. Diagnostic x-ray shielding design—new data and concepts. In: Frey DG, Sprawls P, eds. *The expanding role of medical physics in diagnostic imaging*. Proceedings of the 1997 AAPM Summer School. Madison, WI: Advanced Medical Publishing, 1997.
- Brateman L. The AAPM/RSNA physics tutorial for residents: radiation safety considerations for diagnostic radiology personnel. *RadioGraphics* 1999;19:1037–1055.
- Bushberg JT, Leidholdt EM Jr. Radiation protection. In: Sandler MP et al., eds. *Diagnostic nuclear medicine*, 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2001.
- Dixon RL, Simpkin DJ. New concepts for radiation shielding of medical diagnostic x-ray facilities. In: Frey DG, Sprawls P, eds. *The expanding role of medical physics in diagnostic imaging*. Proceedings of the 1997 AAPM Summer School. Madison, WI: Advanced Medical Publishing, 1997.
- International Commission on Radiological Protection. *Radiological protection and safety in medicine*, vol. 26, no. 2. ICRP no. 73. New York: Pergamon Press, 1996.
- National Council on Radiation Protection and Measurements. *Limitation of exposure to ionizing radiation*. NCRP report no. 116. Bethesda, MD: National Council on Radiation Protection and Measurements, 1993.

- National Council on Radiation Protection and Measurements. *Medical x-ray, electron beam, and gamma-ray protection for energies up to 50 MeV*. NCRP report no. 102. Bethesda, MD: National Council on Radiation Protection and Measurements, 1989.
- National Council on Radiation Protection and Measurements. *Radiation protection in medicine: contemporary issues*. Proceedings of the Thirty-Fifth Annual Meeting NCRP Proceedings no. 21. Bethesda, MD: National Council on Radiation Protection and Measurements, 1999.
- National Council on Radiation Protection and Measurements. *Radiation protection for medical and allied health personnel*. NCRP report no. 105. Bethesda, MD: National Council on Radiation Protection and Measurements, 1989.
- National Council on Radiation Protection and Measurements. *Radiation protection for procedures performed outside the radiology department*. NCRP report no. 133. Bethesda, MD: National Council on Radiation Protection and Measurements, 2000.
- National Council on Radiation Protection and Measurements. *Radionuclide exposure of the embryo/fetus*. NCRP report no. 128. Bethesda, MD: National Council on Radiation Protection and Measurements, 2000.
- National Council on Radiation Protection and Measurements. *Sources and magnitude of occupational and public exposures from nuclear medicine procedures*. NCRP report no. 124. Bethesda, MD: National Council on Radiation Protection and Measurements, 1996.
- Proceedings of the Thirty-third Annual Meeting of the National Council on Radiation Protection and Measurements. The effects of pre- and postconception exposure to radiation. *Teratology* 1999;59(4).
- Title 10 (Energy) of the Code of Federal Regulations (CFR), Part 20: Standards for Protection Against Radiation.
- Title 10 (Energy) of the Code of Federal Regulations (CFR), Part 35: Medical Use of By-product Material.
- Wagner LK, Archer BR. *Minimizing risks from fluoroscopic x-rays: bioeffects, instrumentation, and examination*, 2nd ed. Woodlands, TX: RM Partnership, 1998.

RADIATION DOSIMETRY OF THE PATIENT

The radiation dose to the patient from diagnostic imaging procedures is an important issue and, in the absence of a medical physicist at an institution (e.g., private practice radiology), radiologists are often consulted as the local experts. For x-ray imaging, the doses received by a patient are a function of the imaging modality, the equipment, the technique factors used, and, in the case of fluoroscopy, the ability of the operator to minimize fluoroscopic time. For nuclear medicine procedures, the chemical form of the radiopharmaceutical, its route of administration (e.g., intravenous injection, ingestion, inhalation), the administered activity, the radionuclide, and patient-specific disease states and pharmacokinetics determine the patient dose.

Radiation dosimetry is primarily of interest because radiation dose quantities serve as indices of the risk of biologic damage to the patient. The biologic effects of radiation can be classified as either *deterministic* or *stochastic*. Deterministic effects are believed to be caused by cell killing. If a sufficient number of cells in an organ or tissue are killed, its function can be impaired. Deterministic effects include teratogenic effects to the embryo or fetus, skin damage, and cataracts. For a deterministic effect, a threshold dose can be defined below which the effect will not occur. For doses greater than the threshold dose, the severity of the effect increases with the dose. To assess the likelihood of a deterministic effect on an organ from an imaging procedure, the dose to that organ is estimated.

A stochastic effect is caused by damage to a cell that produces genetically transformed but reproductively viable descendants. Cancer and hereditary effects of radiation are considered to be stochastic. The probability of a stochastic effect, instead of its severity, increases with dose. For stochastic effects, there may not be dose thresholds below which the effects cannot occur.

Radiation dose is sometimes a confusing issue because a simple determination of dose does not tell the whole story, and this is especially true in medical imaging. Radiation dose is defined as the absorbed energy per unit mass, but this says nothing about the total mass of tissue exposed and the distribution of the absorbed energy. Would you prefer to receive a dose of 10 mGy to the whole body or 20 mGy to a finger? The 10-mGy whole-body dose represents about 1,000 times the ionizing energy absorbed for a 70-kg person with a 35-g finger. As another example, a trauma victim has an abdominal computed tomography (CT) scan and receives a radiation dose of 30 mGy to each slice. She then receives a pelvic CT scan, but her radiation dose remains 30 mGy. Only the volume of exposed tissue varies. The most direct way to overcome this conundrum, advocated by some, is to calculate the *energy imparted* (see Chapter 3,

section 3.5). The old term for energy imparted was *integral dose*. The energy imparted is simply the amount of radiation energy absorbed in the body regardless of its distribution. Because dose equals energy divided by mass,

$$\text{Energy imparted (joules)} = \text{Dose (grays)} \times \text{Mass (kilograms)}$$

which is the same as Equation 3-23 in Chapter 3. Calculations of energy imparted can be used to compare radiation doses between different imaging procedures (e.g., a radiographic study versus CT examination). However, a disadvantage is that energy imparted does not account for the different sensitivities of the exposed tissues to biologic damage.

The most common dose measure used for comparing the risk of stochastic effects is the *effective dose*, described in detail in Chapter 3. The effective dose (E) in sieverts (Sv) is calculated from the summation of the products of the dose to each organ (H_T), and its respective weighting factor (w_T , listed in Table 3-5), thus representing the prorated contribution to the total radiation risk from each organ, as was shown in Equation 3-26:

$$E = \sum w_T \times H_T$$

The quantity effective dose has shortcomings. The tissue-weighting factors w_T (Table 3-5) were developed primarily from epidemiologic data and therefore incorporate significant uncertainties. Furthermore, these factors were developed for a reference population of equal numbers of persons of both genders and with a wide range of ages. The quantity effective dose may not apply well to a population that does not match this reference population. For example, if applied to a population of only women, the tissue-weighting factors for the breast and the thyroid would significantly underestimate risk. Another limitation of the quantity effective dose, which is particularly applicable to radiopharmaceutical dosimetry, is that it fails to account for the non-uniform dose distributions within particular organs. Nonetheless, the effective dose is useful for estimating the risk of stochastic effects or comparing radiologic procedures (e.g., a CT body examination versus an F-18 fluorodeoxyglucose positron emission tomography examination) in which the dose distributions to organs are disparate.

Other dosimetric indices are commonly used for comparisons of devices within a particular modality (e.g., comparing doses from CT scanners). The entrance skin exposure (ESE, measured in milliroentgens, mR) or entrance skin dose (in milligray, mGy) is an example of an easily measured parameter that is useful for comparisons between similar imaging procedures, such as screen-film radiography versus computed radiography for chest imaging. Both procedures use the same geometry and similar beam spectra, and therefore the effective dose tracks linearly with the ESE. For example, the effective dose of a 120-kVp, 50-mR ESE chest radiograph would be about three times that of a 120-kVp, 16-mR ESE chest radiographic examination. In this example, the ESE is a reasonable alternative for effective dose, although the absolute values can be very different (e.g., a typical ESE in chest radiography is 20 mR, whereas the effective dose is about 4 mSv). Comparing indices such as the ESE is useful for assessment of equipment performance and calibration, when a comprehensive analysis of effective dose is unnecessary and too complicated to be practical. Several exemplar dose indices, specific to various imaging modalities, are listed in Table 24-1.

To permit comparison of these dose indices among medical institutions, there must be standardized methods for measuring and reporting them. These methods

TABLE 24-1. DOSE INDICES SPECIFIC TO VARIOUS IMAGING MODALITIES

Modality	Dose index	Reference (Chapter no.)
Radiography	Entrance skin exposure (free-in-air)	24
Fluoroscopy	Entrance skin exposure rate (free-in-air)	24
Mammography	Mean glandular dose	8
Nuclear medicine	Activity of radiopharmaceutical injected	2, 18, 19 and 24
Computed tomography	Computed tomographic dose index (CTDI)	13
	Mean scan average dose (MSAD)	

require the use of standard phantoms simulating the portion of the patient being imaged.

In the United States, the Conference of State Radiation Control Program Directors (CRCPD), the state radiologic health regulatory agencies, and the Center for Devices and Radiologic Health of the U.S. Food and Drug Administration (FDA) jointly conduct periodic surveys of the use of x-ray producing devices under the Nationwide Evaluation of X-Ray Trends (NEXT) program. Summaries of NEXT data, published by the CRCPD, such as the example in Table 24-2, provide useful exposure guidelines for comparison of dose indices measured at individual medical institutions. Similar dose guidelines are published by institutions in other countries, including the National Radiological Protection Board of the United Kingdom.

A medical physicist should determine these dose indices on installation of the equipment, annually thereafter, and after any repairs and adjustments that might affect the doses to the patients. These dose indices should be measured using the technique factors commonly used on actual patients at the institution, and they should be compared with appropriate standards, such as NEXT data.

All medical imaging using ionizing radiation involves a compromise between the quality of the images and the radiation exposure to the patient. In the United States, with the exception of mammography, there are no regulatory limits to the amount of radiation received by the patient. (In fluoroscopy the exposure *rate* is regulated but the total fluoroscopic time is not, and therefore the total dose is not.) The physician must decide whether the benefit of the diagnostic procedure justifies the risk to the patient from the radiation exposure. In order to make informed decisions in this regard, referring physicians as well as radiologists must understand the

TABLE 24-2. ENTRANCE SKIN EXPOSURES (ESEs)^a

Projection	Patient thickness (cm)	Grid	SID (cm)	ESE, 200 speed (mR)	ESE, 400 speed (mR)
Abdomen (A/P)	23	Yes	100	490	300
Lumbar spine (A/P)	23	Yes	100	570	330
Full spine (A/P)	23	Yes	183	260	145
Cervical spine (A/P)	13	Yes	100	135	95
Skull (lateral)	15	Yes	100	145	70
Chest (P/A)	23	No	183	15	10
Chest (P/A)	23	Yes	183	25	15

Note: A/P, anteroposterior; P/A, posteroanterior; SID, source-to-image distance.

^aESEs substantially exceeding these values most likely represent excessive patient exposure.

Source: Conference of State Radiation Control Program Directors. *Average patient exposure/dose guides*. Publication 92-4. CRCPD, 1992, Table 1.

TABLE 24-3. ABSORBED DOSES TO SELECTED TISSUES AND EFFECTIVE DOSES FROM SEVERAL COMMON X-RAY EXAMINATIONS IN THE UNITED KINGDOM

Examination	Active bone marrow		Breasts		Uterus (embryo, fetus)		Thyroid		Gonads ^a		Effective dose	
	(mGy)	(mrad)	(mGy)	(mrad)	(mGy)	(mrad)	(mGy)	(mrad)	(mGy)	(mrad)	(mSv)	(mrem)
Chest	0.04	4	0.09	9	*	*	0.02	2	*	*	0.04	4
CT chest	5.9	590	21	2100	0.06	6	2.3	230	0.08, *	8, *	7.8	780
Skull	0.2	20	*	*	*	*	0.4	40	*	*	0.1	10
CT head	2.7	270	0.03	3	*	*	1.9	190	*	*	1.8	180
Abdomen	0.4	40	0.03	3	2.9	290	*	*	2.2, 0.4	220, 40	1.2	120
CT abdomen	5.6	560	0.7	70	8.0	800	0.05	5	8.0, 0.7	800, 70	7.6	760
Thoracic spine	0.7	70	1.3	130	*	*	1.5	150	*	*	1.0	100
Lumbar spine	1.4	140	0.07	7	3.5	350	*	*	4.3, 0.06	430, 6	2.1	210
Pelvis	0.2	20	*	*	1.7	170	*	*	1.2, 4.6	120, 460	1.1	110
CT pelvis	5.6	560	0.03	3	26	2600	*	*	23, 1.7	2300, 170	7.1	710
Intravenous urography	1.9	190	3.9	390	3.6	360	0.4	40	3.6, 4.3	360, 430	4.2	420
Barium enema (including fluoro)	8.2	820	0.7	70	16	1600	0.2	20	16, 3.4	1600, 340	8.7	870
Mammography (film-screen)	*		2	200	*	*	*	*	*	*	0.1	10

Note: *, less than 0.01 mGy (1 mrad); CT, computed tomography.

^aWhen two values are given for the gonads, the first is for the ovaries and the second is for the testes.

Source: Adapted from International Commission on Radiological Protection. *Summary of the current ICRP principles for protection of the patient in diagnostic radiology*, 1993, and data from two publications of the National Radiological Protection Board of the United Kingdom.

risks. The International Commission on Radiological Protection (ICRP) estimates the risk of fatal cancer for exposures to adults of working age to be 4×10^{-2} deaths per Sv (4×10^{-4} per rem). This translates to 1 cancer death per 2,500 people receiving an effective dose of 10 mSv (1 rem). Because of the linear, no-threshold assumption used in risk estimates, risk is presumed to be proportional to the effective dose. For example, there would be a 1 in 25,000 chance that a fatal cancer would result from an effective dose of 1 mSv (0.1 rem), or a 1 in 500 chance of a fatal cancer from an effective dose of 50 mSv (5 rem). The ICRP estimates the risk to be two or three times higher for infants and children, and substantially lower for adults older than 50 years of age (see later discussion of radiation-induced cancer risk, Chapter 25.6).

In addition to understanding the risk per unit effective dose, a working knowledge of the typical dose levels or effective doses for various common radiologic procedures is useful. Table 24-3 is a compilation of typical organ doses and effective doses for a variety of diagnostic imaging procedures. Table 24-4 contains results from a study of patient radiation doses resulting from interventional radiologic procedures.

TABLE 24-4. RANGE OF FLUOROSCOPY SCREENING TIME, FLUOROSCOPY AND RADIOGRAPHY DOSE-AREA PRODUCT VALUES, AND MEAN EFFECTIVE DOSE FOR SELECTED INTERVENTIONAL PROCEDURES

Interventional procedures	Fluoroscopy screening time (min)	Dose-area product (Gy-cm ²) ^a		Mean effective dose (mSv) ^a
		Fluoroscopy	Radiography	
Diagnostic				
AV fistula angiography	1.4–3.8	0.3–36.3	0.8–26.2	0.2
Upper extremity angiography	3.6–8.1	2.5–29.8	3.9–60.2	0.3
Lower extremity angiography	1.8–21.7	1.0–82.2	0.2–1.42	0.8
Nephrostography	0.9–24.7	0.1–72.9	0.0–14.2	2.4
Carotid angiography	2.6–21.0	2.6–79.8	4.7–63.7	4.9
Transjugular hepatic biopsy	2.8–18.7	5.5–95.3	0.0–25.4	5.5
Renal angiography	2.9–7.6	4.7–31.6	12.7–40.2	6.4
Cerebral angiography	2.9–36.0	7.7–121	10.8–76.6	7.4
Femoral angiography	1.8–17.2	1.0–44.6	0.2–91.8	7.5
Thoracic angiography	0.6–114	2.2–167	9.2–115	11.9
PTC	2.9–44.0	12.0–1.92	0.0–11.3	12.8
CT arterial portography	2.3–25.8	17.1–231	0.0–56.5	12.9
Abdominal angiography	1.8–27.1	13.0–102	5.4–227	18.9
Hepatic angiography	3.6–41.8	22.8–168	16.2–149	21.7
Therapeutic				
AV fistula angioplasty	4.8–57.9	0.2–96.3	0.9–45.1	0.3
Biliary stent insertion/removal	0.6–26.3	3.1–137	0.0–18.7	6.9
Nephrostomy	1.3–20.8	0.6–160	0.0–44.7	6.9
Cerebral embolization	15.2–55.8	28.3–61.9	36.3–162	10.5
Renal angioplasty	11.4–27.2	5.4–190	2.6–77.7	13.6
Thoracic therapeutic procedures	4.2–35.0	4.5–182	10.9–168	16.3
Other abdominal therapeutic procedures (excluding hepatic and renal)	6.6–58.8	16.9–413	1.1–229	26.9
TIPS	21.7–100	212–897	61.4–234	83.9

^a1 Gy = 100 rads; 1 mSv = 100 mrem.

Note: AV, arteriovenous; CT, computed tomography; PTC, percutaneous transhepatic cholangiography; TIPS, transjugular intrahepatic portosystemic stent.

Adapted from McParland BJ: A study of patient radiation doses in interventional radiological procedures. *Br J Radiol* 1998;71:175–185.

24.1 X-RAY DOSIMETRY

Medical x-ray imaging procedures such as radiography, fluoroscopy, x-ray CT, and mammography are designed to produce diagnostic image quality with appropriate radiation exposures to the patient. In some cases, it becomes necessary to make patient-specific estimates of the radiation dose. The most common situation that requires accurate estimation of the patient's dose is when a woman has one or more x-ray examinations and later discovers that she is pregnant. Parents often become concerned when their child is exposed to ionizing radiation, and accurate estimation of the dose becomes a way of reassuring them that safe exposure levels were used. Over and above being able to compute radiation doses (which is discussed later), it is necessary to also have the ability to convert radiation doses as quantified in gray or rad into a more meaningful description of risk to the patient, who in most cases is not familiar with dosimetric quantities and units.

Radiographic Procedures

In most cases of having to compute radiation dose, the x-ray procedures have already been performed and one has to reconstruct the situation several weeks or months later. When this happens, some detective work is needed. The patient's medical record should be located and used to assess what examinations were performed. Obtaining the patient's films (or digital images from a picture archiving and communications system [PACS]) is also useful, because this allows an accurate evaluation of where the x-ray fields were positioned. In institutions that have several x-ray rooms, identifying the x-ray machine that was used permits a more accurate estimation of the dose. This may be indicated on the film or determined by asking the technologist who performed the study. For radiographic procedures, it is important to either know or estimate the specific technique factors (peak kilovoltage [kVp] and tube current [milliamperes-seconds, mAs]) for each image acquired. With many newer systems, the technique factors are included with the patient identification data on the film or in the digital record. Most institutions post technique charts that specify the appropriate kVp for a given type of examination, and these are useful for estimating the kVp of a radiographic study if it is not known with certainty. An experienced radiologic technologist can be interviewed to get an estimate of the mAs used for a patient of a given size. Most radiography today uses phototiming (automatic exposure control), whereby the technologist sets the kVp and the x-ray system determines the time of the exposure (and hence the mAs). Information regarding the characteristics (half value level [HVL] or added filtration) of the x-ray machine is also needed for dose assessment. In most cases, these data can be obtained from the annual quality assurance report that should be available from the radiology administration or medical physicist. In other cases, they can be measured after the fact. An estimate of the patient's size is necessary in most situations, because it has a very significant effect on the radiation dose. To summarize, the information that is necessary to make accurate dose estimates from radiographic procedures consists of the following:

- Number of images and radiographic projection of each image (e.g., posteroanterior [PA] and lateral chest views)
- Radiographic technique used for each image (kVp and mAs if available)

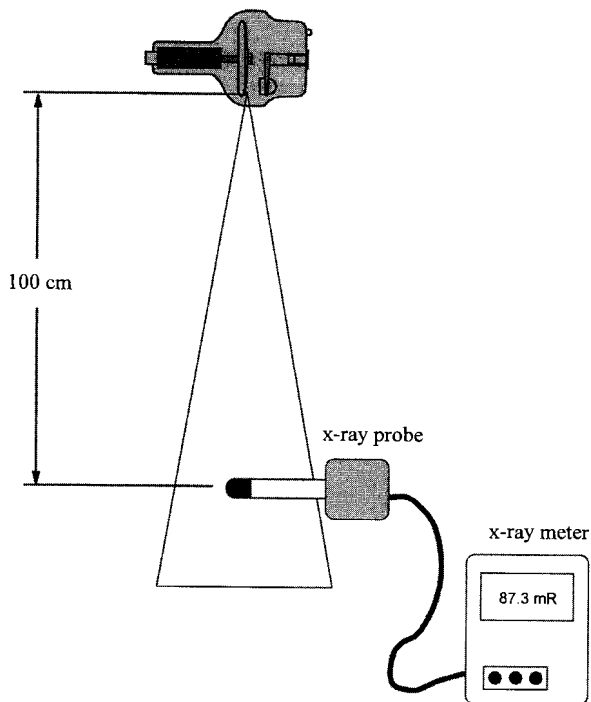


FIGURE 24-1. The geometry for measuring the output free-in-air of a radiographic system is shown.

- Information regarding the characteristics (HVL or inherent filtration) of the beam produced by the x-ray machine
- An estimate of the patient's thickness for each radiographic projection

Figure 24-1 illustrates the measurement geometry that is used to measure the output characteristics of an x-ray tube. The parameter mGy (air kerma) per mAs (or traditional units of mR per mAs), measured at 100 cm from the focal spot, should be obtained for every x-ray tube in an institution, and typical values are shown in Fig. 24-2. Armed with the data in Fig. 24-2 (air kerma per mAs at 100 cm), the

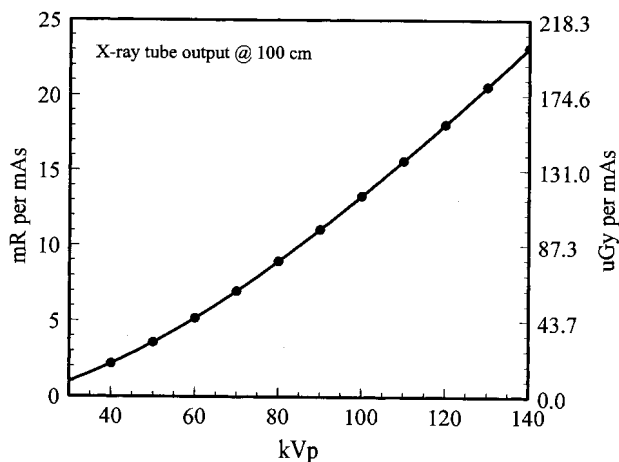


FIGURE 24-2. The output (in mR per mAs and mGy per mAs) for a typical x-ray system measured at a source-to-detector distance of 100 cm.

mAs, and the distance from x-ray source to patient that was used, the air kerma for each examination can be computed, as in the manual technique example given in the next paragraph. Many of the handbooks that are used for dose calculations use the traditional units roentgen and rad. Examples are given using both SI and traditional units.

Manual Technique Example: Calculation of Conceptus Dose

Let us take the example of a woman for whom a single PA abdominal image was acquired using a field size of 35×43 cm (14×17 inches) and a manual technique (i.e., the kVp and mAs were both set by a technologist). Some time after the procedure, the woman discovered that she was pregnant and became concerned about the radiation dose to her fetus. The calculation of fetal dose is facilitated by use of the book *Handbook of Selected Tissue Doses for Projections Common in Diagnostic Radiology* from the Center for Devices and Radiological Health of the U.S. Food and Drug Administration, which uses traditional units. Table 42 from that handbook is reproduced here as Table 24-5. The woman's abdomen was approximately 20 cm thick in the PA projection, and the records indicated that 75 kVp and 25 mAs were used. The HVL of the x-ray beam at 75 kVp was found in the quality assurance report to be 3.0 mm of aluminum, and the tube output at 75 kVp was determined by interpolation of the data in Fig. 24-2 and found to be about 7.9 mR per mAs at 100 cm.

To find the ESE to the patient, the mR per mAs at 100 cm at 75 kVp is multiplied by the mAs and corrected (using the inverse square law) to the distance from the tube focal spot to the front surface of the patient (in this case, 100 cm SID – 20 cm thick patient = 80 cm source to skin distance):

$$\text{ESE} = 25 \text{ mAs} \times \left[\frac{7.9 \text{ mR}}{\text{mAs}} \right] \times \left[\frac{100 \text{ cm}}{80 \text{ cm}} \right]^2 = 309 \text{ mR} = 0.309 \text{ R}$$

The appropriate table from the *Handbook of Selected Tissue Doses for Projections Common in Diagnostic Radiology* is consulted (see Table 24-5). For the 3.0 mm HVL of the beam, the dose to the uterus (site of the exposed fetus) is 173 mrad per 1 R ESE. Using the computed ESE:

TABLE 24-5. TISSUE DOSES FOR 1-R ENTRANCE SKIN EXPOSURE (mRAD) FOR A POSTEROANTERIOR PROJECTION, ABDOMINAL RADIOGRAPHY, 100 CM SOURCE-TO-DETECTOR DISTANCE, FEMALE PATIENT

Location	Half value layer (mm Al)				
	2.0	2.5	3.0	3.5	4.0
Lungs	7.3	10	13	16	19
Active bone marrow	85	114	142	168	192
Thyroid	0.1	0.2	0.3	0.3	0.4
Trunk tissue	115	138	158	175	190
Ovaries	107	151	193	232	266
Uterus	97	136	173	206	235

Source: Adapted from Rosenstein M. *Handbook of selected tissue doses for projections common in diagnostic radiology*. HHS Publication (FDA) 89-8031. U.S. Dept. of Health and Human Services, 1988, Table 42.

$$0.309 \text{ R} \times \frac{173 \text{ mrad}}{\text{R}} = 53.5 \text{ mrad}$$

Therefore, the fetus received a dose of 53.5 mrad (0.53 mGy).

Dose Estimation When Automatic Exposure Control (Phototiming) Is Used

For radiographic rooms where phototiming is used, the mAs of the radiographic examination is generally not known even to the technologist. When the mAs of a radiographic procedure is not known under automatic exposure-controlled acquisition, a different protocol for dose computation should be employed. Typically, slabs of Lucite or other plastic are used to simulate the patient. Over a range of kVp values, various thicknesses of Lucite are placed to simulate the patient. Using automatic exposure control, exposures are made and a table is produced similar to those shown in Table 24-6. The measurement geometry shown in Fig. 24-3 should be used to assess the ESE as a function of kVp and thickness. The exposure meter is located at least 30 cm from the entrant surface of the Lucite to reduce the influence of scattered radiation on the exposure (or air kerma) reading. The exposure value measured using the geometry of Figure 24-3 must be corrected to the surface of the Lucite using the inverse square law. Avoiding the contribution of scatter to the ESE or dose is necessary because most conversion tables are generated under this so-called free-in-air assumption. An example of estimating dose when a phototimed system was used is given in the following paragraph.

TABLE 24-6. ENTRANCE SURFACE DOSE (ESD) AND ENTRANCE SKIN EXPOSURE (ESE) FOR A PHOTOTIMED SYSTEM^a

A: SOURCE-TO-DETECTOR DISTANCE 183 CM (CHEST RADIOGRAPHY).

Measurement	Patient thickness (cm)		
	10 cm	20 cm	30 cm
ESD at 120 kVp (mGy)	0.088	0.724	5.532
ESE at 120 kVp (mR)	10.0	83.0	634

B: SOURCE-TO-DETECTOR DISTANCE 100 CM (STANDARD RADIOGRAPHY).

Peak Kilovoltage (kVp)	Patient thickness					
	10 cm		20 cm		30 cm	
	ESD (mGy)	ESE (mR)	ESD (mGy)	ESE (mR)	ESD (mGy)	ESE (mR)
60	0.166	19.1	2.409	276	—	—
80	0.129	14.8	1.498	172	16.177	1,853
100	0.110	12.6	1.121	128	10.751	1,231
120	0.099	11.3	0.921	105	8.146	933

^aThese values were calculated for a 200-speed computed radiography system operating with a grid. (A 200-speed system is calibrated to receive about 8.73 μGy [1 mR] to the detector.) A computer-generated spectral model was used, and the patient was modeled as water. The scatter-to-primary ratio (~2–3) exiting the patient and reaching the detector was assumed to offset the Bucky factor of the grid (~3–4), and therefore the calculation dealt with only primary radiation.

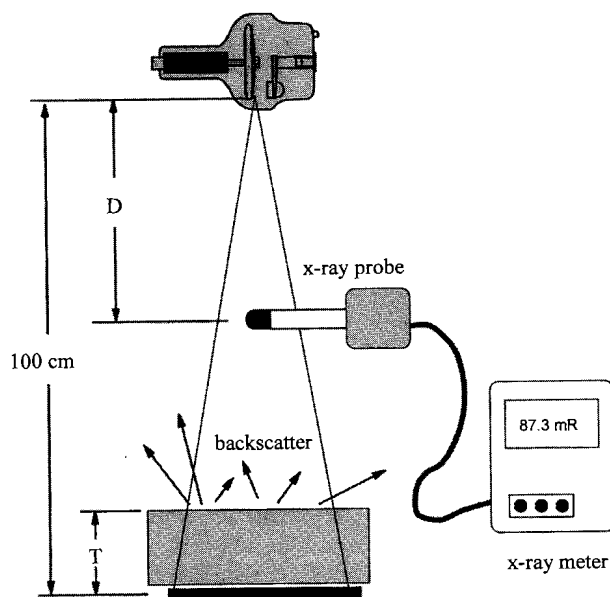


FIGURE 24-3. The geometry for determining the thickness-dependent surface dose under phototiming (automatic exposure control) conditions is illustrated. The x-ray probe needs to be positioned away from the front surface of the Lucite to avoid measuring backscattered radiation. The dose measurements should then be corrected to the surface of the Lucite by the inverse square law.

Calculation of Effective Dose to the Patient under AEC

A 15-year-old patient received a PA chest radiograph, and her parents are concerned about the radiation dose. The technologist used automatic exposure control and, although the kVp was known to be 120 kVp, the mAs used in the examination was not recorded. The patient was about 20 cm thick (PA); however, due to the low density of lungs her effective thickness was estimated to be slightly less, about 15 cm thick. The image was acquired using a 200-speed computed radiography system.

For rooms that use phototiming routinely (i.e., most modern radiographic suites), the quality assurance report should, as mentioned, include the determination of entrance dose as a function of patient thickness (using Lucite) and kVp. Table 24-6 illustrates this data. For 120 kVp and a patient thickness of about 15 cm, the entrance dose in the chest geometry (Table 24-6A, 183 cm source-to-detector distance) must be interpolated to 15 cm from the 10 cm and 20 cm data given. Interpolation of entrance dose across thickness should be performed by logarithmic interpolation, since x-ray attenuation increases exponentially with patient thickness. Because 15 cm is exactly halfway between 10 cm and 20 cm, the interpolation involves simply averaging the logarithms of the entrance dose and computing the exponential of this average.

$$e^{\left[\frac{\ln(0.88) + \ln(0.724)}{2} \right]} = 0.252 \text{ mGy}$$

This equation demonstrates that the estimated entrance dose to a 15-cm patient is about 0.252 mGy.

To calculate the effective dose to the patient based on an entrance dose of 0.252 mGy, we use tables in *Estimation of Effective Dose in Diagnostic Radiology from*

TABLE 24-7. EFFECTIVE DOSE PER ENTRANCE SURFACE DOSE (mSV PER mGY) FOR CHEST RADIOGRAPHY 183 CM SOURCE-TO-DETECTOR DISTANCE

X-ray potential (kVp)	Filtration (mm of Al)	Anteroposterior (mSv/mGy)	Posteroanterior (mSv/mGy)	Lateral (mSv/mGy)
90	2	0.176	0.116	0.074
90	3	0.196	0.131	0.084
90	4	0.210	0.143	0.091
100	2	0.190	0.128	0.081
100	3	0.208	0.143	0.091
100	4	0.222	0.155	0.098
110	2	0.201	0.139	0.088
110	3	0.219	0.154	0.097
110	4	0.232	0.165	0.104
120	2	0.211	0.149	0.094
120	3	0.228	0.163	0.103
120	4	0.240	0.174	0.110

Source: From *Estimation of effective dose in diagnostic radiology from entrance surface dose and dose-area product measurements*. National Radiation Protection Board of Great Britain, 1994, Table 7.

Entrance Surface Dose and Dose-Area Product Measurements, NRPB Report R262 (1994), published by the National Radiological Protection Board of the United Kingdom (partially reproduced here as Table 24-7). For a 120-kVp beam with 3 mm of added aluminum filtration, the conversion factor for the PA chest projection is 0.163 mSv per mGy (effective dose per entrance dose):

$$\text{Effective dose} = 0.252 \text{ mGy} \times \frac{0.163 \text{ mSv}}{\text{mGy}} = 0.041 \text{ mSv} = 41 \text{ } \mu\text{Sv}$$

Therefore, the effective dose to this girl is estimated to be 41 μSv . To put this into perspective, the average effective dose from natural background radiation in the United States is approximately 3 mSv per year (including radon, see Chapter 23.1). Thus, the radiation received from the chest radiograph was about 1.4% of the child's annual background radiation, comparable to the background radiation effective dose accumulated over 5 days of living in the United States.

24.2 RADIOPHARMACEUTICAL DOSIMETRY: THE MIRD METHOD

The Medical Internal Radiation Dosimetry (MIRD) Committee of the Society of Nuclear Medicine has developed a methodology for calculating the radiation dose to selected organs and the whole body from internally administered radionuclides. The MIRD formalism takes into account variables associated with the deposition of ionizing radiation energy and those associated with the biologic system for which the dose is being calculated. Although some of the variables are known with a relatively high degree of accuracy, others are best estimates or simplified assumptions that, taken together, provide an estimate of the dose to the average (reference) adult, adolescent, child, and fetus.

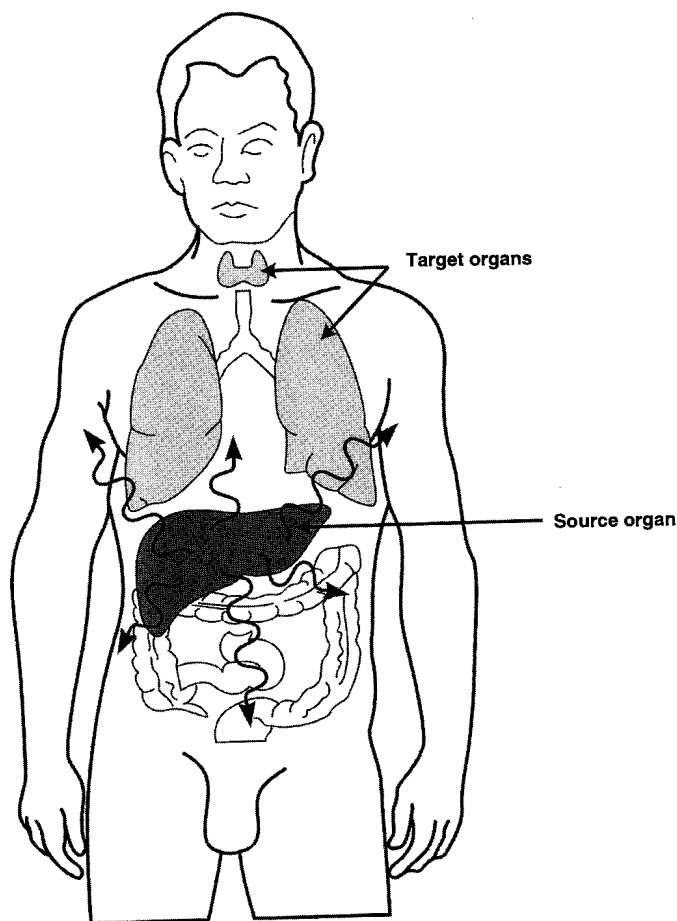


FIGURE 24-4. Illustration of source and target organ concept for calculation of the dose to the lungs and thyroid gland (target organs) from a radiopharmaceutical (e.g., technetium-99m sulfur colloid) primarily located in the liver (source organ). Note that the relative geometry and mass of the source and target organs together with the physical decay characteristic of the radionuclide and its associated radiopharmaceutical kinetics all play a part in determining the dose to any particular organ.

The MIRD formalism has two elements: (a) estimation of the quantity of radiopharmaceutical in various “source” organs and (b) estimation of the radiation absorbed in selected “target” organs from the radioactivity in source organs (Fig. 24-4.)

MIRD Formalism

The MIRD formalism designates the source organ in which the activity is located, r_h , and the target organ for which the dose is being calculated, r_k . The mean dose to a particular target organ ($\bar{D}_{r_k}$) is calculated from the following relationship:

$$\bar{D}_{r_k} = \sum_h \bar{A}_h S(r_k \leftarrow r_h) \quad [24-1]$$

where $\bar{A}_h$ is the cumulated activity (in $\mu\text{Ci}\cdot\text{hr}$) for each source organ (r_h) and S (expressed in $\text{rads}/\mu\text{Ci}\cdot\text{hr}$) is the absorbed dose in the target, r_k , per unit of

cumulated activity in each source organ. When summed for each source organ $\left(\sum_b\right)$, the mean dose to a specific target organ is $(\bar{D}_{rk})$, and the units are $\sum_b \bar{A} (\mu\text{Ci-hr}) \cdot S(r_k \leftarrow r_b) (\text{rads}/\mu\text{Ci-hr}) = \text{rads}$. The MIRD Committee has not yet published S factors in SI units; however, this conversion is anticipated.

Cumulated Activity

The cumulated activity is the total number of disintegrations from the radionuclide located in a particular source organ. The units of $\mu\text{Ci-hr}$ express the total number of disintegrations:

$$\text{Disintegrations/Time} \times \text{Time} = \text{Disintegrations}$$

The cumulated activity depends on (a) the portion of the injected dose “taken up” by the source organ (A_f) and (b) the rate of elimination from the source organ.

Figure 24-5 shows a simple kinetics model representing the accumulation and elimination of radioactivity in a source organ. If we assume that a fraction (f) of the injected activity (A_0) is localized in the source organ and there is exponential biologic excretion, two processes will act to reduce the total activity in the source organ: (a) physical decay of the radionuclide, as represented by its physical half-life T_p , and (b) biologic elimination, as represented by the biologic half-life T_b . These two factors work together to produce an effective half-life (T_e) that is calculated as follows:

$$T_e = \frac{T_p \cdot T_b}{T_p + T_b} \quad [24-2]$$

The cumulated activity, defined as the total number of disintegrations occurring in the source organ, is equal to the area under the time-activity curve. In a manner

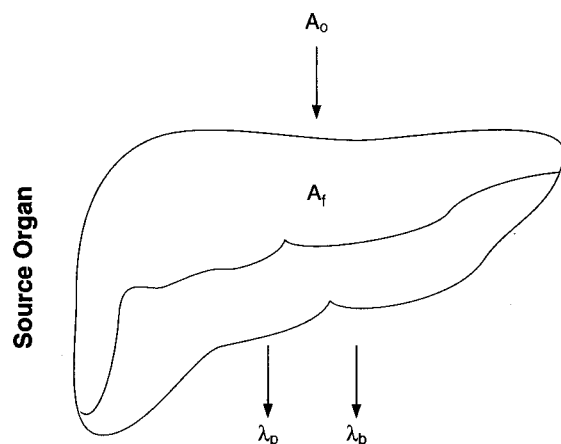


FIGURE 24-5. Simplified kinetics model of the accumulation and elimination of radioactivity in a source organ (e.g., liver). A fraction (f) of the injected activity (A_0) is localized in the source organ in which the initial activity (A_t) is reduced by physical decay and biologic excretion of the radiopharmaceutical.

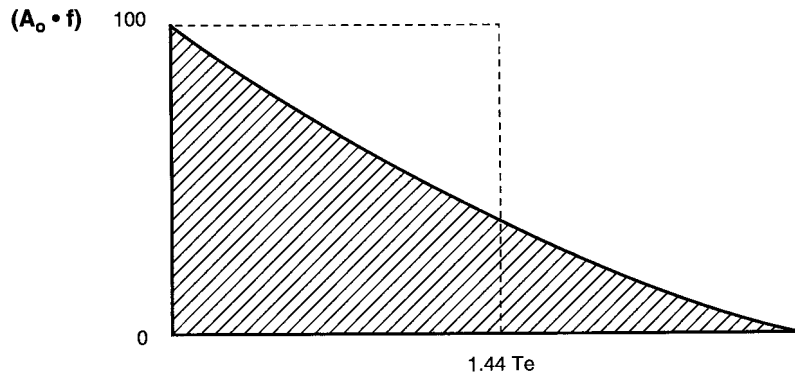


FIGURE 24-6. The cumulated activity is equal to the area under the curve, which is numerically equal to $(A_0 \cdot f)(1.44 T_e)$.

analogous to radioactive decay, the activity remaining in a source organ after a time (t) is

$$A_t = fA_0e^{-\lambda_e t} \quad [24-3]$$

where λ_e is the effective decay constant, equal to $0.693/T_e$.

The area under this curve (Fig. 24-6) can be shown to be equal to the product of the initial amount of activity in the source organ ($f \cdot A_0$) and the so-called “average life” (τ), which is calculated from the following relationship:

$$\tau = \frac{T_e}{\ln 2} = \frac{T_e}{0.693} = 1.44 T_e \quad [24-4]$$

Therefore, the cumulated activity $\tilde{A}_b$ can be expressed (in $\mu\text{Ci}\cdot\text{hr}$) as the product of these terms:

$$\tilde{A}_b = A_0 \cdot f_b \cdot 1.44 T_e \quad [24-5]$$

where A_0 is expressed in μCi and T_e in hours.

S Factor

The S factor is the dose to the target organ per unit of cumulated activity in a specified source organ (i.e., $\text{rad}/\mu\text{Ci}\cdot\text{hr}$). It is determined by a number of factors, including (a) the mass of the target organ, (b) the type and amount of ionizing radiation emitted per disintegration, and (c) the fraction of the emitted radiation energy that reaches and is absorbed by the target organ.

Each S factor is specific to a particular source organ/target organ combination. S factors are provided in tabular form for most common diagnostic and therapeutic radionuclides. Table 24-8 lists some S factors for technetium-99m (Tc-99m). The MIRD Committee has not yet published S factors in SI units; however, this conversion is anticipated. Table 24-9 summarizes the variables associated with the MIRD method of internal dosimetry calculation.

TABLE 24-8. Tc-99m S FACTORS FOR SOME SOURCE/TARGET ORGAN COMBINATIONS

Target organs (r_k)	Source organs (r_h)									
	Adrenals	Bladder contents	Intestinal Tract				Kidneys	Liver	Lungs	Other tissue (muscle)
			Stomach contents	SI contents	ULI contents	LLI contents				
Adrenals	3.1E-03	1.5E-07	2.7E-06	1.0E-06	9.1E-07	3.6E-07	1.1E-05	4.5E-06	2.7E-06	1.4E-06
Bladder wall	1.3E-07	1.6E-04	2.7E-07	2.6E-06	2.2E-06	6.9E-06	2.8E-07	1.6E-07	3.6E-08	1.8E-06
Bone (total)	2.0E-06	9.2E-07	9.0E-07	1.3E-06	1.1E-06	1.6E-06	1.4E-06	1.1E-06	1.5E-06	9.8E-07
GI (stomach wall)	2.9E-06	2.7E-07	1.3E-04	3.7E-06	3.8E-06	1.8E-06	3.6E-06	1.9E-06	1.8E-06	1.3E-06
GI (SI)	8.3E-07	3.0E-06	2.7E-06	7.8E-05	1.7E-05	9.4E-06	2.9E-06	1.6E-06	1.9E-07	1.5E-06
GI (ULI Wall)	9.3E-07	2.2E-06	3.5E-06	2.4E-05	1.3E-04	4.2E-06	2.9E-06	2.5E-06	2.2E-07	1.6E-06
GI (LLI Wall)	2.2E-07	7.4E-06	1.2E-06	7.3E-06	3.2E-06	1.9E-04	7.2E-07	2.3E-07	7.1E-08	1.7E-06
Kidneys	1.1E-05	2.6E-07	3.5E-06	3.2E-06	2.8E-06	8.6E-07	1.9E-04	3.9E-05	8.4E-07	1.3E-06
Liver	4.9E-06	1.7E-07	2.0E-06	1.8E-06	2.6E-06	2.5E-07	3.9E-06	4.6E-05	2.5E-06	1.1E-06
Lungs	2.4E-06	2.8E-08	1.7E-06	2.2E-07	2.6E-07	7.9E-08	8.5E-07	2.5E-06	5.2E-05	1.3E-06
Marrow (red)	3.6E-06	2.2E-06	1.6E-06	4.3E-06	3.7E-06	5.1E-06	3.8E-06	1.6E-06	1.9E-06	2.0E-06
OTH TISS (muscle)	1.4E-06	1.8E-06	1.4E-06	1.5E-06	1.5E-06	1.7E-06	1.3E-06	1.1E-06	1.3E-06	2.7E-06
Ovaries	6.1E-07	7.3E-06	5.0E-07	1.1E-05	1.2E-05	1.8E-05	1.1E-06	4.5E-07	9.4E-08	2.0E-06
Pancreas	9.0E-06	2.3E-07	1.8E-05	2.1E-06	2.3E-06	7.4E-07	6.6E-06	4.2E-06	2.6E-06	1.8E-06
Skin	5.1E-07	5.5E-07	4.4E-07	4.1E-07	4.1E-07	4.8E-07	5.3E-07	4.9E-07	5.3E-07	7.2E-07
Spleen	6.3E-06	6.6E-07	1.0E-05	1.5E-06	1.4E-06	8.0E-07	8.6E-07	9.2E-07	2.3E-06	1.4E-06
Testes	3.2E-08	4.7E-06	5.1E-08	3.1E-07	2.7E-07	1.8E-06	8.8E-08	6.2E-08	7.9E-09	1.1E-06
Thyroid	1.3E-07	2.1E-09	8.7E-08	1.5E-08	1.6E-08	5.4E-09	4.8E-08	1.5E-07	9.2E-07	1.3E-06
Uterus (nongravid)	1.1E-06	1.6E-05	7.7E-07	4.6E-06	5.4E-06	7.1E-06	9.4E-07	3.9E-07	8.2E-08	2.3E-06
Total body	2.2E-06	1.9E-06	1.9E-06	2.4E-06	2.2E-06	2.3E-06	2.2E-06	2.2E-06	2.0E-06	1.9E-06

Note: GI, gastrointestinal; SI, small intestine; ULI, upper large intestine; LLI, lower large intestine; OTH TISS, other tissue. Bold italicized correspond to values in MIRD example problem.

Source: Medical Internal Radiation Dosimetry (MIRD) Committee of the Society of Nuclear Medicine.

TABLE 24-9. VARIABLES IN THE MIRD SYSTEM OF INTERNAL DOSIMETRY CALCULATION

Symbol	Quantity	Describes	Equivalence	Units
T_e	Effective half-life	Time required for the activity in the source organ to decrease by one-half	$T_p \cdot T_b / T_p + T_b$	hours (hr)
A_o	Activity	Administered activity	—	μCi
f_h	Fractional uptake	Fraction of the administered activity (A_o) in source organ (h)	—	0 to 1
$\tilde{A}_h$	Cumulated activity	Total number of disintegrations in source organ h	$A_o \cdot f_h \cdot 1.44 T_e$	$\mu\text{Ci-hr}$
$S(r_k \leftarrow r_h)$	"S" factor	Dose in target organ (r_k) per unit of cumulated activity ($\tilde{A}_h$) in source organ (r_h)	—	$\text{rad}/\mu\text{Ci-hr}$
$\bar{D}_{r_k}$	Mean target organ dose	Radiation energy deposited per unit mass of the target organ (r_k)	$\sum \tilde{A}_h S(r_k \leftarrow r_h)$	rad

Example of MIRD Dose Calculation

The simplest example of a MIRD internal dose calculation would be the situation in which the radiopharmaceutical instantly localizes in a single organ and remains there with an infinite T_b . Although most radiopharmaceutical dose calculations do not meet both of these assumptions, this hypothetical example is considered for the purpose of demonstrating the MIRD technique.

Example: A patient is injected with 3 mCi of Tc-99m-sulfur colloid. Estimate the radiation absorbed dose to the (a) liver, (b) testes, (c) red bone marrow, and (d) total body.

Assumptions:

1. All of the injected activity is uniformly distributed in the liver.
2. The uptake of Tc-99m-sulfur colloid in the liver from the blood is instantaneous.
3. There is no biologic removal of Tc-99m-sulfur colloid from the liver.

In this case, the testes, red bone marrow, and total body are target organs, whereas the liver is both a source and a target organ. The average dose (D) to any target organ (r_k) can be estimated by applying the simplified MIRD formalism in Equation 24-1.

$$\bar{D}_{r_k} = \sum_h \tilde{A}_h S(r_k \leftarrow r_h)$$

Step 1: Calculate the cumulated activity ($\tilde{A}_h$) in the source organ (i.e., liver) by inserting the appropriate values into Equation 24-5:

$$\tilde{A}_h = A_o \cdot f_h \cdot 1.44 \cdot T_e$$

$$\tilde{A}_h = 3,000 \mu\text{Ci} \times 1 \times 1.44 \times 6.02 \text{ hr} = 2.60 \times 10^4 \mu\text{Ci-hr}$$

Note that because T_b is infinite, $T_e = T_p$.

Step 2: Find the appropriate S factors for each target/source combination and the radionuclide of interest (i.e., Tc-99m) from Table 24-8. The appropriate S factors are found at the intersection of the source organ's column (i.e., liver) and the individual target organ's row (i.e., liver, testes, red marrow or total body).

Target (r_k)	Source (r_h)	S Factor
Liver	Liver	4.6×10^{-5}
Red bone marrow	Liver	1.6×10^{-6}
Testes	Liver	6.2×10^{-8}
Total body	Liver	2.2×10^{-6}

Step 3: Organize the assembled information in a table of organ doses:

Target organ (r_k)	$\bar{A}_h$ ($\mu\text{Ci-hr}$)	$\times$	$S(r_k \leftarrow r_h)$ ($\text{rad}/\mu\text{Ci-hr}$)	$=$	$\bar{D}_{r_k}$ (rad)
Liver	2.60×10^4		4.6×10^{-5}		1.2
Testes	2.60×10^4		6.2×10^{-8}		0.002
Red bone marrow	2.60×10^4		1.6×10^{-6}		0.042
Total body	2.60×10^4		2.2×10^{-6}		0.057

Note that the relatively small dose to target organs other than the liver is mostly the result of only a small fraction of the isotropically emitted penetrating radiation (i.e., gamma rays) being absorbed by the other organ systems (e.g., testes, red bone marrow).

Accuracy of Dose Calculations

Although the MIRD method provides reasonable estimates of organ doses, the typical application of this technique usually includes several significant assumptions, limitations, and simplifications that, taken together, could result in differences of as much as 50% between the true and calculated doses. These include the following:

1. The radioactivity is assumed to be uniformly distributed in each source organ. This is rarely the case and, in fact, significant pathology (e.g., cirrhosis of the liver) or characteristics of the radiopharmaceutical may result in a highly nonuniform activity distribution.
2. The organ sizes and geometries are idealized into simplified shapes to aid mathematical computations.
3. Each organ is assumed to be homogeneous in density and composition.
4. The phantoms for the "reference" adult, adolescent, and child are only approximations of the physical dimensions of any given individual.
5. Although the radiobiologic effect of the dose occurs at the molecular level, the energy deposition is averaged over the entire mass of the target organs and therefore does not reflect the actual microdosimetry on a molecular or cellular level.
6. Dose contributions from bremsstrahlung and other minor radiation sources are ignored.
7. With a few exceptions, low-energy photons and all particulate radiations are assumed to be absorbed locally (i.e., nonpenetrating).

In addition, the time integral activity ($\bar{A}_h$) used for each initial organ dose estimate during the developmental phase of a radiopharmaceutical is usually based on laboratory animal data. This information is only slowly adjusted by human investigation and quantitative evaluation of biodistributions and kinetics. The FDA does

not currently require manufacturers to update their package inserts as better radiopharmaceutical dosimetry becomes available. Therefore, organ doses listed in package inserts (especially those of older agents) are often an unreliable source of dosimetry information.

In addition to the development of child and fetal dosimetry models, advances in the use of radioimmunotherapeutic pharmaceuticals has increased the need for patient-specific dosimetry that takes advantage of individual kinetics and anatomic information. A good overview of these issues and the MIRD model can be found in several recent reviews (see Suggested Reading).

For the most commonly used diagnostic and therapeutic radiopharmaceutical agents, Appendix D.2 summarizes the typically administered adult dose, the organ receiving the highest radiation dose and its dose, the gonadal dose, and the adult effective dose. For most of these same radiopharmaceuticals, Appendix D.3 provides a table of effective doses per unit activity administered in 15-, 10-, 5-, and 1-year-old patients, and Appendix D.4 provides a table of absorbed doses to the embryo or fetus per unit activity administered to the mother at early, 3, 6, and 9 months gestation.

SUGGESTED READING

- Loevinger R, Budinger T, Watson E. *MIRD primer for absorbed dose calculations*. New York: Society of Nuclear Medicine, 1991.
- Parry RA, Glaze SA, Archer BR. The AAPM/RSNA physics tutorial for residents. Typical patient radiation doses in diagnostic radiology. *Radiographics* 1999;19:1289–1302.
- Rosenstein M. *Handbook of selected tissue doses for projections common in diagnostic radiology*. Rockville, MD: U.S. Food and Drug Administration, Center for Devices and Radiological Health, 1988.
- Simpkin DJ. Radiation interactions and internal dosimetry in nuclear medicine. *Radiographics* 1999;19:155–167.
- Toohy RE, Stabin MG, Watson EE. The AAPM/RSNA physics tutorial for residents. Internal radiation dosimetry: principles and applications. *Radiographics* 2000;20:533–546.
- Wagner LK, Lester RG, Saldana LR. *Exposure of the pregnant patient to diagnostic radiations: a guide to medical management*, 2nd ed. Madison, WI: Medical Physics Publishing, 1997.
- Zanzonico PB. Internal radionuclide radiation dosimetry: a review of basic concepts and recent developments. *J Nuclear Med* 2000;41:297–308.

RADIATION BIOLOGY

Rarely have beneficial applications and hazards to human health followed a major scientific discovery more rapidly than with the discovery of ionizing radiation. Shortly after the discovery of x-rays in 1895 and natural radioactivity in 1896, biologic effects from ionizing radiation were being observed. Two months after their discovery, x-rays were being used to treat breast cancer. Unfortunately, the development and implementation of radiation protection techniques lagged behind the rapidly increasing use of radiation sources. Within the first 6 months of their use, several cases of erythema, dermatitis, and alopecia were reported among x-ray operators and their patients. The first report of a skin cancer ascribed to x-rays was in 1902, to be followed 8 years later by experimental confirmation. However, it was not until 1915 that the first radiation protection recommendations were made by the British Roentgen Society, followed by similar recommendations from the American Roentgen Ray Society in 1922.

The study of the action of ionizing radiation on healthy and diseased tissue began the scientific discipline known as radiation biology. Radiation biologists seek to understand the sequence of events that occurs after the absorption of energy from ionizing radiation, the damage that is produced, and the mechanisms that exist to compensate for or repair the damage. A century of radiobiologic research has amassed more information about the effects of ionizing radiation on living systems than about almost any other physical or chemical agent.

This chapter reviews the consequences of ionizing radiation exposure, beginning with the chemical basis on which radiation damage is initiated and its subsequent effects on cells, organ systems, and the whole body. This is followed by a review of the concepts and risks associated with radiation-induced carcinogenesis, genetic effects, and the special concerns regarding radiation exposure to the fetus.

Determinants of the Biologic Effects of Radiation

There are many factors that determine the biologic response to radiation exposure. In general, these factors include variables associated with the radiation source and the system being irradiated. The identification of these biologic effects depends on the method of observation (Table 25-1). Radiation-related factors include the dose, type, and energy of the radiation as well as the dose rate and the conditions under which the dose is delivered. The radiosensitivity and complexity of the biologic system determine the type of response from a given exposure. In general, complex organisms exhibit more sophisticated repair mechanisms. Responses seen at the

TABLE 25-1. DETERMINANTS OF BIOLOGIC DAMAGE FROM IONIZING RADIATION

Radiation	System	Observational variables
Quality	Molecular	Time
Quantity	Cellular vs. multicellular	Macroscopic vs. microscopic
Dose rate	organism	Structural vs. functional changes
Exposure conditions	Plant vs. animal vs. human	

molecular or cellular level may or may not result in adverse clinical effects in humans. Furthermore, although some responses to radiation exposure appear instantaneously, others take weeks, years, or even decades to appear.

Classification of Biologic Effects

Biologic effects of radiation exposure can be classified as either *stochastic* or *deterministic*. A stochastic effect is one in which the probability of the effect, rather than its severity, increases with dose. Radiation-induced cancer and genetic effects are stochastic in nature. For example, the probability of radiation-induced leukemia is substantially greater after an exposure to 1 Gy (100 rad) than to 0.01 Gy (1 rad), but there will be no difference in the severity of the disease if it occurs. Stochastic effects are believed not to have a dose threshold, because injury to a few cells or even a single cell could theoretically result in production of the disease. Therefore even minor exposures may carry some, albeit small, increased risk. It is the basic assumption that *risks increase with dose and there is no threshold dose below which risks cease to exist* that is the basis of modern radiation protection programs, the goal of which is to keep exposures *as low as reasonably achievable* (ALARA). Stochastic effects are regarded as the principal health risk from low-dose radiation, including exposures in the diagnostic radiology and nuclear medicine department.

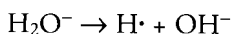
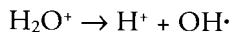
If a radiation exposure is very high, the predominant biologic effect is cell killing that results in degenerative changes in the exposed tissue. In this case, the severity of the injury, rather than its probability of occurrence, increases with dose. These so-called deterministic effects differ from stochastic effects in that they require much higher doses to produce an effect. There is also a *threshold* dose below which the effect is not seen. Cataracts, erythema, fibrosis, and hematopoietic damage are some of the deterministic effects that can result from large radiation exposures. Many of these effects are discussed in the sections entitled "Response of Organ Systems to Radiation" and "Acute Radiation Syndrome." Deterministic effects can be caused by serious radiation accidents and can be observed in healthy tissue that is unavoidably irradiated during radiation therapy. However, with the exception of some lengthy, fluoroscopically guided interventional procedures (see Chapter 23), they are unlikely to occur as a result of diagnostic imaging procedures or routine occupational exposure.

25.1 INTERACTION OF RADIATION WITH TISSUE

Energy is deposited randomly and rapidly (in less than 10^{-10} sec) by ionizing radiation via excitation, ionization, and thermal heating. One of the fundamental

tenets in radiation biology is that radiation-induced injury to an organism always begins with chemical changes at the atomic and molecular level. For example, observable effects such as chromosome breakage, cell death, oncogenic transformation, and acute radiation sickness, all have their origin in radiation-induced chemical changes to important biomolecules. The changes produced in molecules, cells, tissues, and organs are not unique and cannot be distinguished from damage produced by other physical or chemical agents. The biologic changes resulting from the radiation-induced damage are apparent only after a period of time (latent period) that varies from minutes to weeks and even years depending on the biologic system and the initial dose. Only a fraction of the radiation energy deposited brings about chemical changes; the vast majority of the energy is deposited as heat. The heat produced is of little biologic significance compared with the heat generated by normal metabolic processes. For example, it would take more than 1,000 Gy (100,000 rad), a supralethal dose, to raise the temperature of tissue by 1°C.

Radiation interactions that produce biologic changes are classified as either *direct* or *indirect*. The change takes place by direct action if a biologic macromolecule such as DNA, RNA, or protein becomes ionized or excited by an ionizing particle or photon passing through or near it. Indirect effects are the result of radiation interactions within the medium (e.g., cytoplasm) which create reactive chemical species that in turn interact with the target molecule. Because 70% to 85% of the mass of living systems is composed of water, the vast majority of radiation-induced damage from medical irradiation is mediated through indirect action on water molecules. The absorption of radiation by a water molecule results in an ion pair (H_2O^+ , H_2O^-). The H_2O^+ ion is produced by the ionization of H_2O , whereas the H_2O^- ion is produced via capture of a free electron by a water molecule. These ions are very unstable; each dissociates to form another ion and a *free radical*:



Free radicals, symbolized by a dot on the right-hand side of the chemical symbol, are atomic or molecular species that have unpaired orbital electrons. The H^+ and OH^- ions do not typically produce significant biologic damage because of their extremely short lifetimes ($\sim 10^{-10}$ sec) and their tendency to recombine to form water. Free radicals are extremely reactive chemical species that can undergo a variety of chemical reactions. Free radicals can combine with other free radicals to form nonreactive chemical species such as water (e.g., $\text{H}\cdot + \text{OH}\cdot = \text{H}_2\text{O}$), in which case no biologic damage occurs, or with each other to form other molecules such as hydrogen peroxide (e.g., $\text{OH}\cdot + \text{OH}\cdot = \text{H}_2\text{O}_2$), which are highly toxic to the cell. Free radicals can act as strong oxidizing or reducing agents by combining directly with macromolecules. The damaging effect of free radicals is enhanced by the presence of oxygen. Oxygen stabilizes free radicals and reduces the probability of free radical recombination to form water. Oxygen combines with the hydrogen radical to form the highly reactive hydroperoxyl radical (e.g., $\text{H}\cdot + \text{O}_2 = \text{HO}_2\cdot$). Although their lifetimes are limited (less than 10^{-5} sec), free radicals can diffuse in the cell, producing damage at locations remote from their origin. Free radicals may inactivate cellular mechanisms directly or via damage to genetic material (DNA and RNA), and they are believed to be the primary cause of biologic damage from low linear energy transfer (LET) radiation.

Repair mechanisms exist within cells that are capable, in many cases, of returning the cell to its preirradiated state. For example, if a break occurs in a single strand of DNA, the site of the damage may be identified and the break may be repaired by rejoining the broken ends. There are specific endonucleases and exonucleases that are capable of repairing damage to the DNA. If there is base damage on a single strand, enzymatic excision occurs, and the intact complementary strand of the DNA molecule provides the template on which to reconstruct the correct base sequence. If the damage is too severe or these repair mechanisms are compromised or overwhelmed by excessive radiation exposure, the cell will be transformed. The clinical consequence of such a transformation depends on a number of variables. For example, if the damage were to the DNA at a location that prevented the cell from producing albumin, the clinical consequences would be insignificant considering the number of cells remaining with the ability to produce this serum protein. If, however, the damage were to the DNA at a location that was responsible for controlling the rate of cell division, the clinical consequences could be the formation of a tumor or cancer. Heavily irradiated cells, however, often die during mitosis, thus preventing the propagation of seriously defective cells. Figure 25-1 summarizes the physical and biologic responses to ionizing radiation.

Experiments with cells and animals have shown that the biologic effect of radiation depends not only on factors such as the dose, dose rate, environmental con-

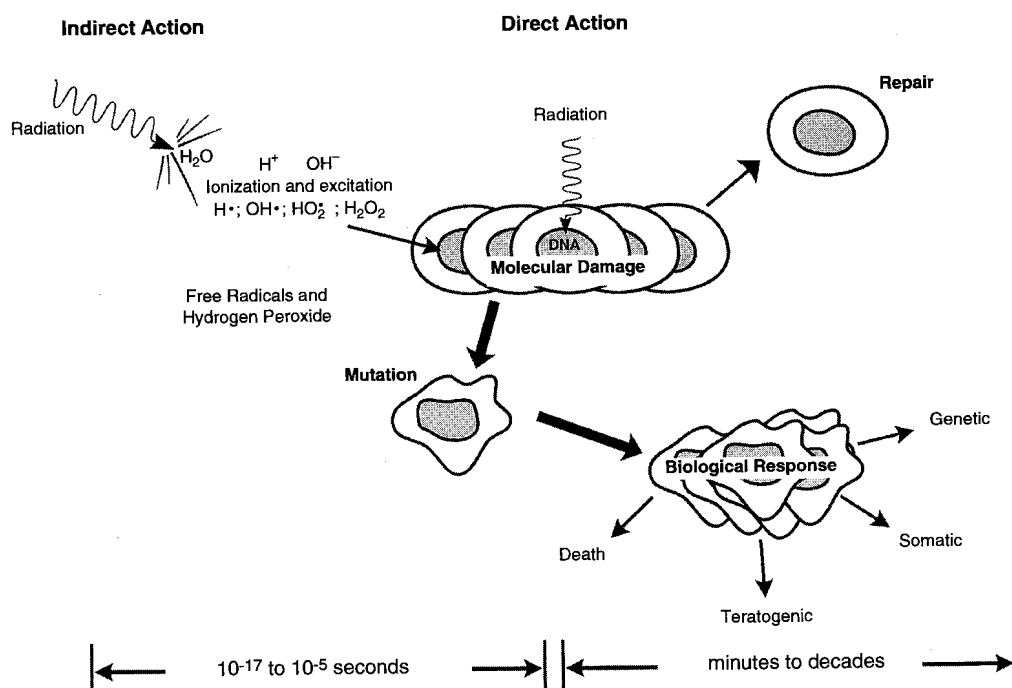


FIGURE 25-1. Physical and biologic responses to ionizing radiation. Ionizing radiation causes damage either directly by damaging the molecular target or indirectly by ionizing water, which in turn generates free radicals that attack molecular targets. The physical steps that lead to energy deposition and free radical formation occur within 10^{-5} to 10^{-6} seconds, whereas the biologic expression of the physical damage may occur seconds or decades later.

dition at the time of irradiation, and radiosensitivity of the biologic system but also on the spatial distribution of the energy deposition. The LET is a parameter that describes the average energy deposition per unit path length of the incident radiation (see Chapter 3). Although all ionizing radiations are capable of producing the same types of biologic effects, the magnitude of the effect per unit dose differs. Equal doses of radiation of different LETs do not produce the same biologic response. To evaluate the effectiveness of different types of radiations and their associated LETs, experiments are performed that compare the dose required of the test radiation to produce the same specific biologic response produced by a particular dose of a reference radiation (typically x-rays produced by a potential of 250 kVp). The term relating the effectiveness of the test radiation to the reference radiation is called the *relative biological effectiveness* (RBE). The RBE is defined, for identical exposure conditions, as follows:

$$\text{RBE} = \frac{\text{Dose of 250-kVp x-rays required to produce effect X}}{\text{Dose of test radiation required to produce effect X}}$$

The RBE is initially proportional to LET: As the LET of the radiation increases, so does the RBE (Fig. 25-2). The increase is attributed to the higher spe-

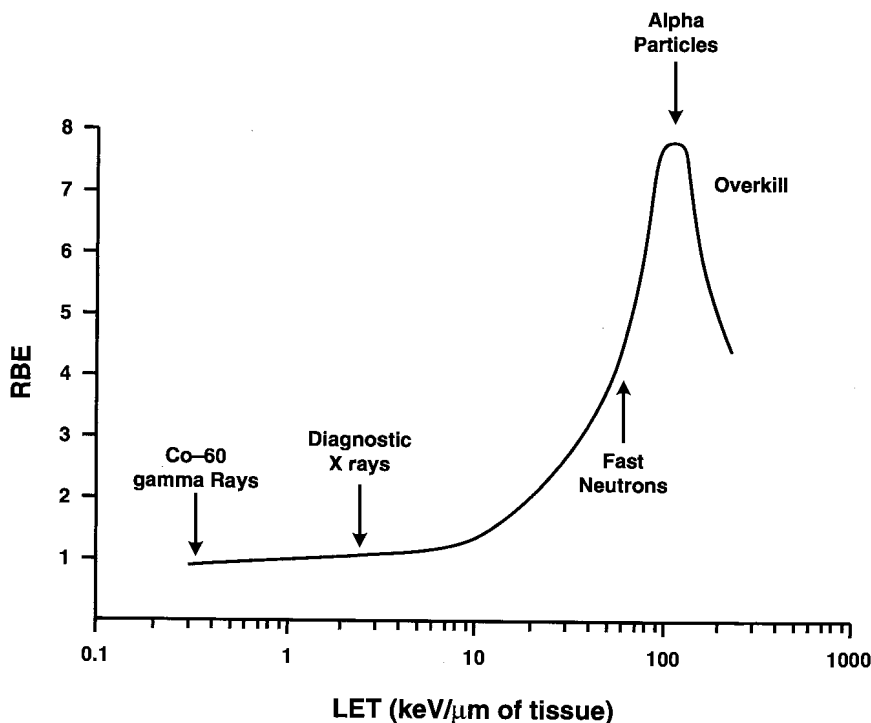


FIGURE 25-2. The relative biological effectiveness (RBE) of a given radiation is an empirically derived term that, in general, all other factors being held constant, increases with the linear energy transfer (LET) of the radiation. However, beyond approximately 100 keV/μm, the radiation becomes less efficient due to overkill (i.e., the maximal potential damage has already been reached), and the increase in LET beyond this point results in wasted dose. For example, at 500 keV/μm many of the cells may sustain three or more ionizing events, when only two are required to kill the cell.

cific ionization (i.e., ionization density) associated with high-LET radiation (e.g., alpha particles) and its relative advantage in producing cellular damage compared with low-LET radiation (e.g., x-rays, gamma rays). However, beyond 100 keV/ μm in tissue, the RBE decreases with increasing LET, because of the overkill effect. Overkill refers to the deposition of radiation in excess of that necessary to kill the cell. The RBE ranges from less than 1 to more than 20. For a particular type of radiation, the RBE depends on the biologic end point. For example, chromosomal mutation, cataract formation, or acute lethality of test animals may be used as an end point. The RBE also depends on the total dose and the dose rate. Despite these deficiencies, the RBE is a useful radiobiologic tool that helps to characterize the potential damage from various types of radiation. The RBE is an essential element in establishing the radiation weighting factors (w_R) discussed in Chapter 3.

25.2 CELLULAR RADIOBIOLOGY

Cellular Targets

Although the critical lesions responsible for cell killing have not been identified, it has been established that the radiation-sensitive targets are located in the nucleus and not the cytoplasm of the cell. Cells contain numerous macromolecules, only some of which are essential for cell survival. For example, there are many copies of various enzymes within a cell; the loss of one particular copy would not significantly affect the cell's function or survival. However, if a key molecule, for which the cell has no replacement (e.g., DNA), is damaged or destroyed, the result may be cell death. There is considerable evidence that damage to DNA is the primary cause of radiation-induced cell death. This concept of key or critical targets has led to a model of radiation-induced cellular damage called *target theory* in which critical targets can be inactivated by one or more ionization events (called *hits*).

Radiation Effects on DNA

The deposition of energy (directly or indirectly) by ionizing radiation induces chemical changes in large molecules that may then undergo a variety of structural changes. These structural changes include (a) hydrogen bond breakage, (b) molecular degradation or breakage, and (c) intermolecular and intramolecular cross-linking. The rupture of the hydrogen bonds that link base pairs in DNA may lead to irreversible changes in the secondary and tertiary structure of the macromolecule that compromise genetic transcription and translation. Molecular breakages also may involve DNA. They may occur as single-strand breaks, double-strand breaks (in which both strands of the double helix break simultaneously at approximately the same nucleotide pair), base loss, or base changes. Single-strand breakage between the sugar and the phosphate can rejoin provided there is no opportunity for the broken portion of the strands to separate. The presence of oxygen potentiates the damage by causing the end of the broken strand to become peroxidized, thus preventing it from rejoining. A double-strand break can occur if two single-strand breaks are juxtaposed or when a single, densely ionizing particle (e.g., an alpha particle) produces a break in both strands. DNA double-strand breaks are very genotoxic lesions that can result in chromosome aberrations. The genomic instability resulting from persistent or incorrectly repaired double-strand breaks can

lead to carcinogenesis through activation of oncogenes, inactivation of tumor suppressor genes, or loss of heterozygosity. Single-strand breaks are more easily repaired than double-strand breaks and are more likely to result from the sparse ionization pattern that is characteristic of low-LET radiation. Figure 25-3 illustrates some of the common forms of damage to DNA.

Molecular cross-linking is another common macromolecular structural change. Cross-linking is believed to result from the formation of reactive sites at the point of chain breakage. DNA can undergo cross-linking between two DNA molecules, between DNA and a protein, or between two base pairs within the DNA double helix (e.g., thymidine dimerization induced by ultraviolet radiation). Regardless of its severity or consequences, the loss or change of a base is considered a type of *mutation*.

Although mutations can have serious implications, changes in the DNA are discrete and do not necessarily result in structural changes in the chromosomes. However, chromosome breaks produced by radiation do occur and can be observed microscopically during anaphase and metaphase, when the chromosomes are shortened. Radiation-induced chromosomal lesions can occur in both somatic and germ cells and, if not repaired before DNA synthesis, may be transmitted during mitosis and meiosis. Chromosomal damage that occurs before DNA replication is referred to as *chromosome aberrations*, whereas that occurring after DNA synthesis is called

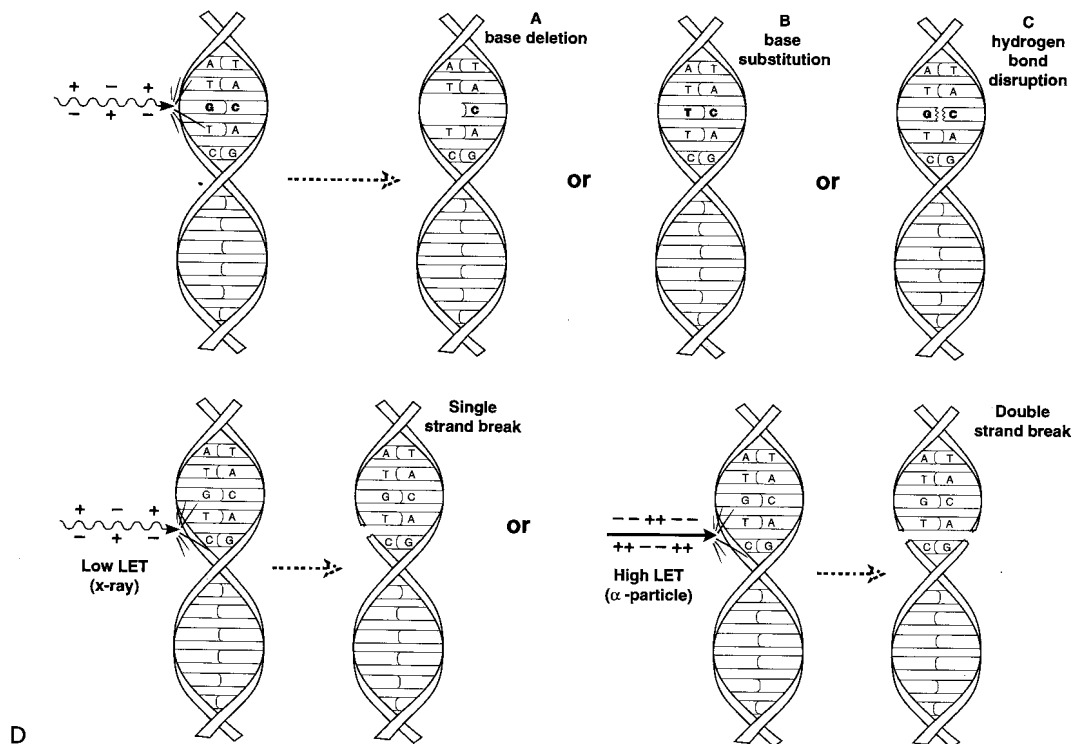


FIGURE 25-3. Several examples of DNA mutations. **A:** Base deletion. **B:** Base substitution. **C:** Hydrogen bond disruption. **D:** Single- and double-strand breaks.

chromatid aberrations. Unlike chromosomal aberrations, only one of the daughter cells will be affected if only one of the chromatids of a pair is damaged.

Repair of chromosomal aberrations depends on several factors, including the stage of the cell cycle and the type and location of the lesion. There is a strong force of cohesion between broken ends of chromatic material. Interchromosomal and intrachromosomal recombination may occur in a variety of ways, yielding many types of aberrations such as rings and dicentrics. Figure 25-4 illustrates some of the more common chromosomal aberrations and the consequences of replication and

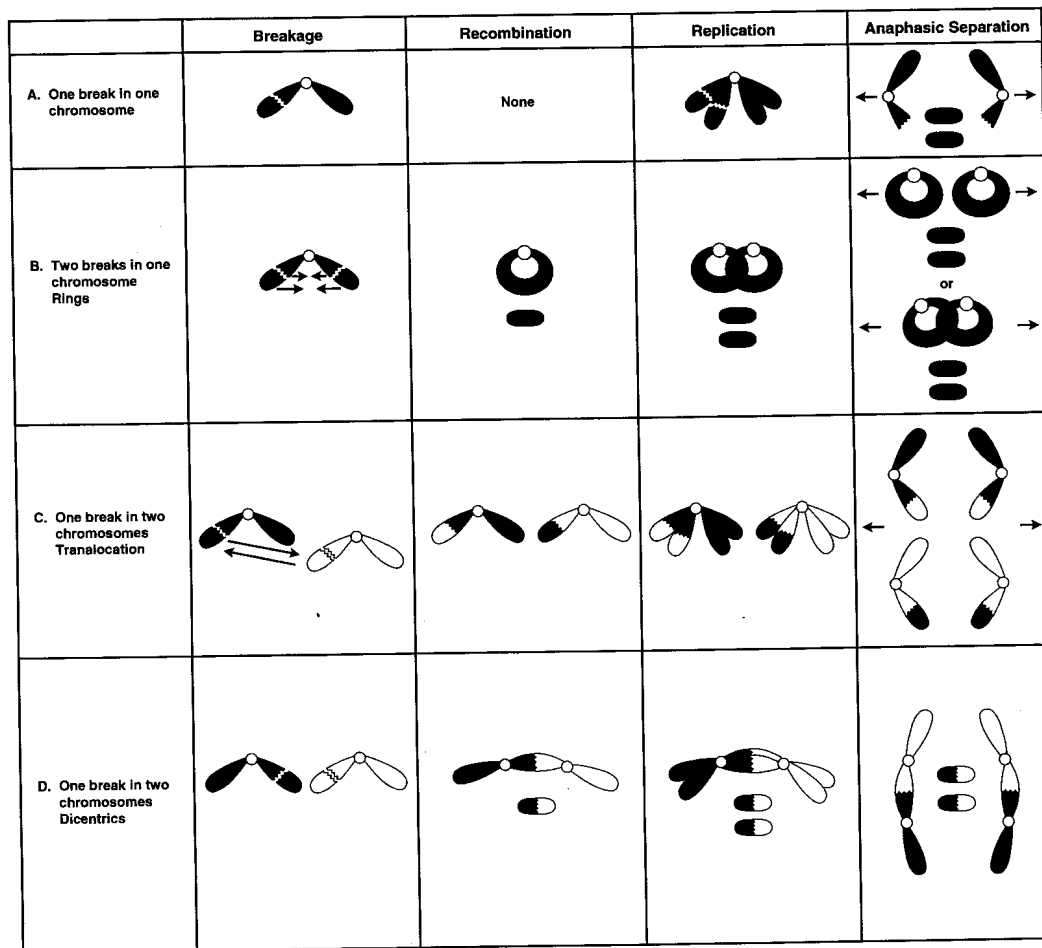


FIGURE 25-4. Several examples of chromosomal aberrations and the effect of recombinations, replication, and anaphasic separation. **A:** Single break in one chromosome, which results in centric and acentric fragments. The acentric fragments are unable to migrate and are transmitted to only one of the daughter cells. These fragments are eventually lost in subsequent divisions. **B:** Ring formation may result from two breaks in the same chromosome in which the two broken ends of the centric fragment recombine. The ring-shaped chromosome undergoes normal replication, and the two (ring-shaped) sister chromatids separate normally at anaphase—unless the centric fragment twists before recombination, in which case the sister chromatids will be interlocked and unable to separate. **C:** Translocation may occur when two chromosomes break and the acentric fragment of one chromosome combines with the centric fragment of the other and vice versa, or **D:** the two centric fragments recombine with each other at their broken ends, resulting in the production of a dicentric.

anaphasic separation. The extent of the total genetic damage transmitted with chromosomal aberrations depends on a variety of factors, such as the cell type, the number and kind of genes deleted, and whether the lesion occurred in a somatic or in a gametic cell.

Chromosomal aberrations are known to occur spontaneously. The scoring of chromosomal aberrations in human lymphocytes can be used as a biologic dosimeter to estimate the dose of radiation received after an accidental exposure. T lymphocytes are cultured from a sample of the patient's blood and then stimulated to divide, allowing a karyotype to be performed. The cells are arrested at metaphase, and the frequency of rings and dicentrics is scored. Total body doses in excess of 0.25 Gy (25 rad) can be detected in this way. Although these chromosomal aberrations are gradually lost from circulation, they do persist for a considerable period after the exposure and can still be detected in the survivors of the atomic bombs in Hiroshima and Nagasaki, more than 50 years later.

Cellular Radiosensitivity

Although a host of biologic effects are induced by radiation exposure, the study of radiation-induced cell death (often defined as the loss of reproductive integrity) is particularly useful in assessing the relative biologic impact of various types of radiation and exposure conditions. The use of reproductive integrity as a biologic effects marker is somewhat limited, however, in that it is applicable only to proliferating cell systems (e.g., stem cells). For differentiated cells that no longer have the capacity for cell division (e.g., muscle and nerve cells), cell death is often defined as loss of specific metabolic functions.

Cell Survival Curves

Cells grown in tissue culture that are lethally irradiated may fail to show evidence of morphologic changes for long periods; however, reproductive failure eventually occurs. The most direct method of evaluating the ability of a single cell to proliferate is to wait until enough cell divisions have occurred to form a visible colony. Counting the number of *colonies* that arise from a known number of individual cells irradiated *in vitro* and cultured provides a way to determine easily the relative radiosensitivity of particular cell lines, the effectiveness of different types of radiation, or the effect of various environmental conditions. The loss of the ability to form colonies as a function of radiation exposure can be described by cell survival curves.

Survival curves are usually presented with the radiation dose plotted on a linear scale on the x-axis and the surviving fraction of cells plotted on a logarithmic scale on the y-axis. The response to radiation is defined by three parameters: the extrapolation number (n), the quasithreshold dose (D_q), and the D_0 dose (Fig. 25-5). The extrapolation number is found by extrapolating the linear portion of the curve back to its intersection with the y-axis. This number is thought to represent either the number of targets (i.e., critical molecular sites) in a cell that must be "hit" once by a radiation event to inactivate the cell or the number of "hits" required on a single critical target to inactivate the cell. The extrapolation number for mammalian cells ranges between 2 and 10. The D_0 describes the radiosensitivity of the cell population under study. The D_0 dose is the reciprocal of the slope of the linear

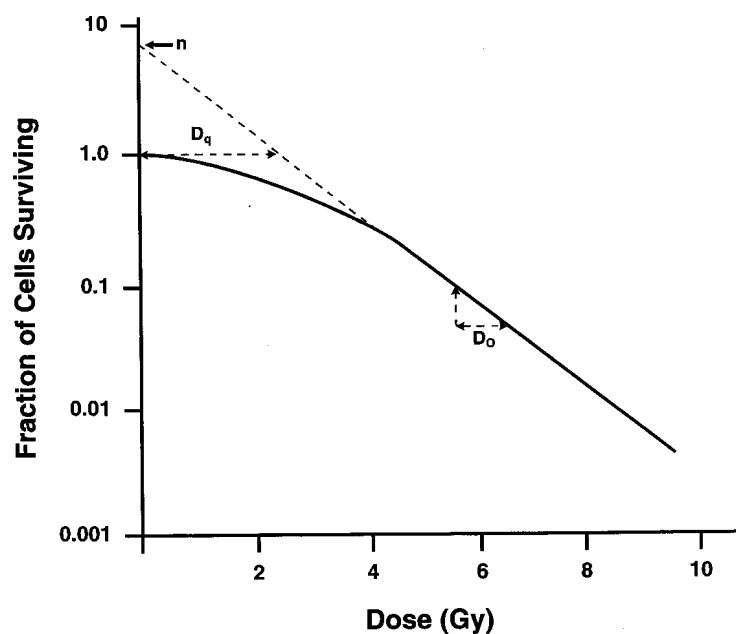


FIGURE 25-5. Typical cell survival curve illustrating the portions of the curve used to derive the extrapolation number (n), the quasithreshold dose (D_q), and the D_0 dose.

portion of the survival curve, and it is the dose of radiation that produces, along the linear portion of the curve, a 37% reduction in the number of viable cells. Radiore-sistant cells have a higher D_0 than radiosensitive cells do. A lower D_0 implies less survival per dose. The D_0 for mammalian cells ranges from approximately 1 to 2 Gy (100 to 200 rad).

In the case of low-LET radiation, the survival curve is characterized by an initial “shoulder” before the linear portion of the semilogarithmic plot. D_q defines the width of the shoulder region of the cell survival curve and is a measure of sublethal damage. Sublethal damage is a concept derived from target theory that predicts that irradiated cells remain viable until enough hits have been received to inactivate a critical target (or targets). The presence of the shoulder in a cell survival curve provides clear evidence that, for low-LET radiation, damage produced by a single radiation interaction with a cell’s critical target or targets is insufficient to produce reproductive death.

Factors Affecting Cellular Radiosensitivity

Cellular radiosensitivity can be influenced by a variety of factors that can enhance or diminish the response to radiation or alter the temporal relationship between the exposure and a given response. These factors can be classified as either *conditional* or *inherent*. Conditional radiosensitivities are those physical or chemical factors that exist before and/or at the time of irradiation. Some of the more important conditional factors affecting dose-response relationships are discussed in the following

paragraphs, including dose rate, LET, and the presence of oxygen. Inherent radiosensitivity includes those biologic factors that are characteristics of the cells themselves, such as the mitotic rate, the degree of differentiation, and the stage of the cell cycle.

Conditional Factors

The rate at which a dose of low-LET radiation is delivered has been shown to affect the degree of biologic damage for a number of biologic end points including chromosomal aberrations, reproductive failure, and death. In general, high dose rates are more effective at producing biologic damage than low dose rates. The primary explanation for this effect is the diminished potential for repair of radiation damage at high dose rates. Cells have a greater opportunity to repair sublethal damage at low dose rates than at higher dose rates, reducing the amount of damage and increasing the survival fraction. Figure 25-6 shows an example of the dose rate effect on cell survival. Note that the broader shoulder and higher D_0 associated with low-dose-rate exposure indicate its diminished effectiveness compared with the same dose delivered at a higher dose rate. This dose-rate effect is not seen with high-LET radiation primarily because the dense ionization tracks typically produce enough hits in critical targets to overwhelm the cell's repair mechanism or produce types of damage that cannot be repaired. However, for a given dose rate, high-LET radiation is considerably more effective in producing cell damage than low-LET radiation (Fig. 25-7).

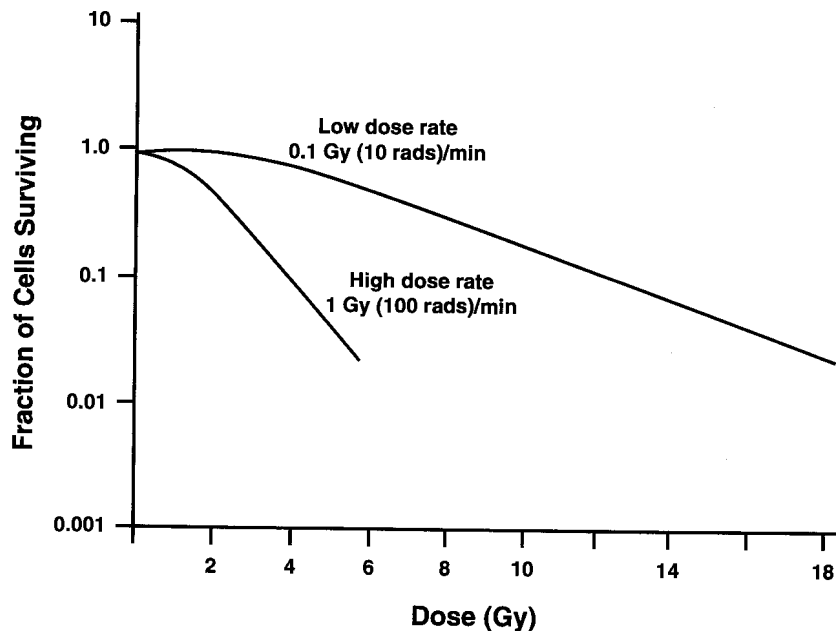


FIGURE 25-6. Cell survival curves illustrating the effect of dose rate. Lethality is reduced because repair of sublethal damage is enhanced when a given dose of radiation is delivered at a low versus a high dose rate.

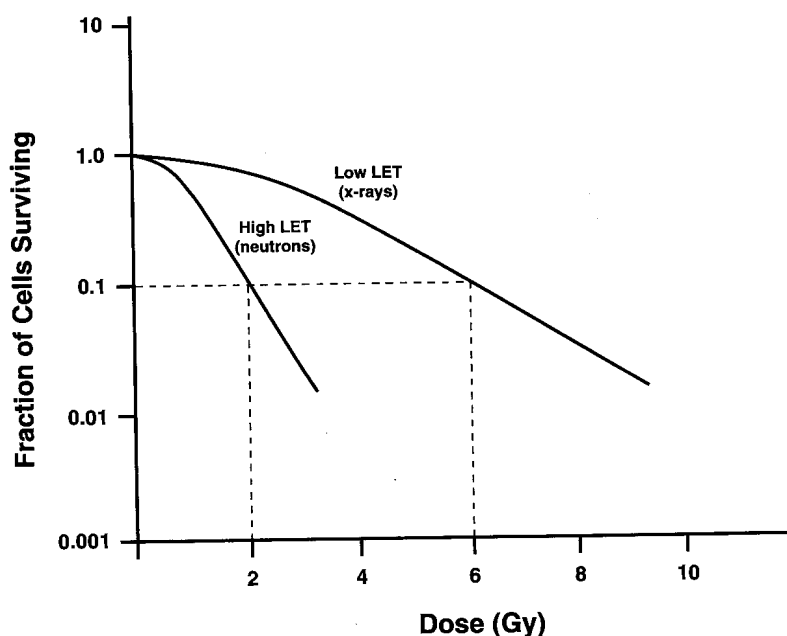


FIGURE 25-7. Cell survival curves illustrating the greater damage produced by radiation with high linear energy transfer (LET) at 10% survival. High-LET radiation is three times as effective as the same dose of low-LET radiation in this example.

For a given radiation dose, a reduction in radiation damage is also observed when the dose is *fractionated* over a period of time. This technique is fundamental to radiation therapy for cancer. The intervals between doses (typically days or weeks) allow the repair mechanisms in healthy tissue to gain an advantage over the tumor by repairing some of the sublethal damage. Figure 25-8 shows an idealized experiment in which a dose of 1,000 Gy is delivered either all at once or in five fractions of 200 Gy with sufficient time between fractions for repair of sublethal damage. For low-LET radiation, the decreasing slope of the survival curve with decreasing dose rate (see Fig. 25-6) and the reoccurrence of the shoulder with fractionation (see Fig. 25-8) are clear evidence of repair.

The presence of free oxygen increases the damage caused by low-LET radiation by inhibiting the recombination of free radicals to form harmless chemical species and by inhibiting the repair of damage caused by free radicals. This effect is demonstrated in Fig. 25-9, which shows a cell line irradiated under aerated and hypoxic conditions. The relative effectiveness of the radiation to produce damage at various oxygen tensions is described by the oxygen enhancement ratio (OER). The OER is defined as the dose of radiation that produces a given biologic response in the absence of oxygen divided by the dose of radiation that produces the same biologic response in the presence of oxygen. Increasing the oxygen concentration at the time of irradiation has been shown to enhance the killing of otherwise hypoxic (i.e., radioresistant) neoplastic cells. The OER for mammalian cells in cultures is typically between 2 and 3 for *low*-LET radiation. High-LET damage is not primarily

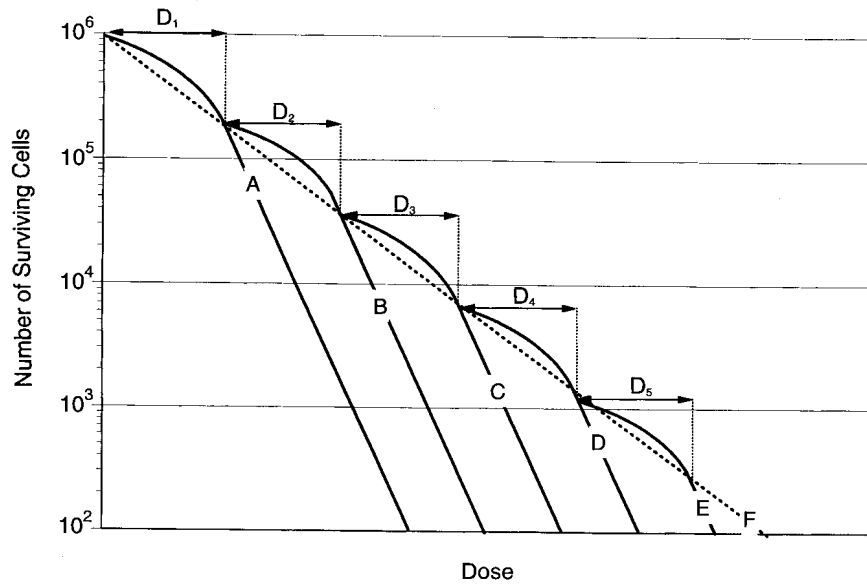


FIGURE 25-8. Idealized fractionation experiment depicting the survival of a population of 10^6 cells as a function of dose. Curve A represents one fraction of 10 Gy (1,000 rads). Curves F represent the same total dose as in curve A delivered in equal fractionated doses (D_1 through D_5) of 2 Gy (200 rad) each, with intervals between fractions sufficient to allow for repair of sublethal damage. (Modified from Hall EJ. *Radiobiology for the radiologist*, 5th ed. Philadelphia: Lippincott Williams & Wilkins, 2000.)

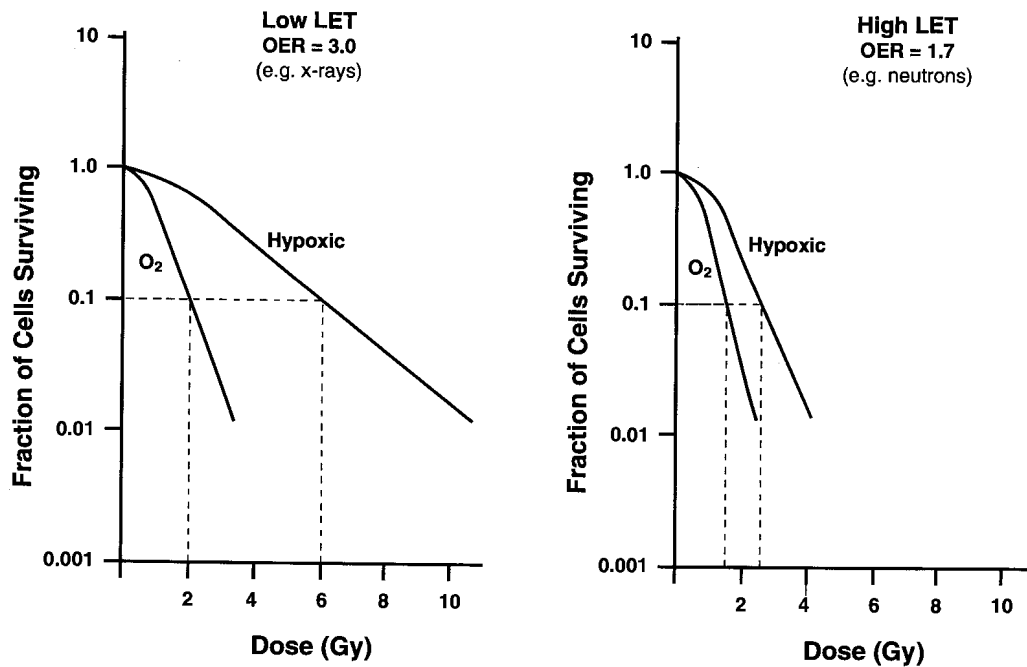


FIGURE 25-9. Cell survival curves demonstrating the effect of oxygen during high and low linear energy transfer (LET) irradiation.

mediated through free radical production, and therefore the OER for high-LET radiation can be as low as 1.0.

Inherent Factors

In 1906, two French scientists, J. Bergonié and L. Tribondeau, performed a series of experiments that evaluated the relative radiosensitivity of rodent germ cells at different stages of spermatogenesis. From these experiments, some of the fundamental characteristics of cells that affect their relative radiosensitivities were established. The Law of Bergonié and Tribondeau states that radiosensitivity is greatest for those cells that (a) have a high mitotic rate, (b) have a long mitotic future, and (c) are undifferentiated. With only a few exceptions (e.g., lymphocytes), this law provides a reasonable characterization of the relative radiosensitivity of cells *in vitro* and *in vivo*. For example, the pluripotential stem cells in the bone marrow have a high mitotic rate, a long mitotic future, are poorly differentiated, and are extremely radiosensitive compared with other cells in the body. On the other end of the spectrum, the fixed postmitotic cells that comprise the central nervous system (CNS) are relatively radioresistant. This classification scheme was refined in 1968 by Rubin and Casarett, who defined five cell types according to characteristics that affect their radiosensitivity (Table 25-2).

The stage of the cells in the reproductive cycle at the time of irradiation greatly affects their radiosensitivity. Figure 25-10 shows the phases of the cell

TABLE 25-2. CLASSIFICATION OF CELLULAR RADIOSENSITIVITY

Cell type	Characteristics	Examples	Radiosensitivity
VIM	Rapidly dividing; undifferentiated; do not differentiate between divisions	Type A spermatogonia Erythroblasts Crypt cells of intestines Basal cells of epidermis	Most radiosensitive
DIM	Actively dividing; more differentiated than VIMs; differentiate between divisions	Intermediate spermatogonia Myelocytes	Relatively radiosensitive
MCT	Irregularly dividing; more differentiated than VIMs or DIMs	Endothelial cells Fibroblasts	Intermediate in radiosensitivity
RPM	Do not normally divide but retain capability of division; differentiated	Parenchymal cells of the liver Lymphocytes ^a	Relatively radioresistant
FPM	Do not divide; differentiated	Some nerve cells Muscle cells Erythrocytes (RBCs) Spermatozoa	Most radioresistant

Note: VIM, vegetative intermitotic cells; DIM, differentiating intermitotic cells; MCT, multipotential connective tissue cells; RPM, reverting postmitotic cells; FPM: fixed postmitotic cells.

^aLymphocytes, although classified as relatively radioresistant by their characteristics, are in fact very radiosensitive.

Source: Rubin P, Casarett GW. Clinical radiation pathology as applied to curative radiotherapy. *Clin Pathol Radiat* 1968;22:767-768.

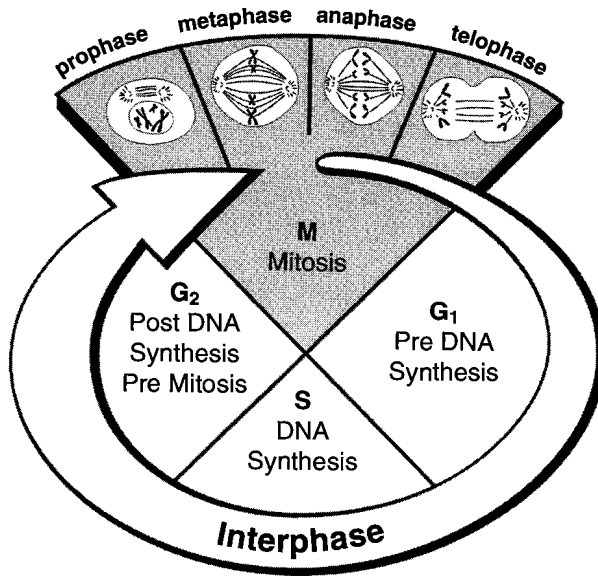


FIGURE 25-10. Phases of the cell's reproductive cycle. The time between cell divisions is called *interphase*. Interphase includes the period after mitosis but before DNA synthesis (G_1), which is the most variable of the phases; followed by *S phase*, during which DNA synthesis occurs; followed by RNA synthesis in G_2 , all leading up to mitosis (*M phase*), the events of which are differentiated in *prophase*, *metaphase*, *anaphase*, and *telophase*. (Adapted from Bushong SC. *Radiologic science for technologists*, 3rd ed. St. Louis: CV Mosby, 1984.)

reproductive cycle. Experiments indicate that, in general, cells are most sensitive to radiation during mitosis (M phase) and RNA synthesis (G_2), less sensitive during the preparatory period for DNA synthesis (G_1), and least sensitive during DNA synthesis (S phase).

25.3 RESPONSE OF ORGAN SYSTEMS TO RADIATION

The response of an organ system to radiation depend not only on the dose, dose rate, and LET of the radiation but also on the relative radiosensitivities of the cells that comprise both the functional parenchyma and the supportive stroma. In this case, the response is measured in terms of morphologic and functional changes of the organ system as a whole rather than simply changes in cell survival kinetics.

The response of an organ system after irradiation occurs over a period of time whose onset and period of expression are inversely proportional to the dose. The higher the dose, the shorter the interval before the physiologic manifestations of the damage become apparent (latent period) and the shorter the period of expression during which the full extent of the radiation-induced damage is evidenced. There are practical threshold doses below which no significant changes are apparent. In most cases, the pathology induced by radiation is undistinguishable from naturally occurring pathology.

Regeneration and Repair

Healing of tissue damage produced by radiation occurs by means of cellular *regeneration* and *repair* (Fig. 25-11). Regeneration refers to replacement of the damaged cells in the organ by cells of the same type, thus replacing the lost functional capacity. Repair refers to the replacement of the damaged cells by fibrotic scar tissue, in

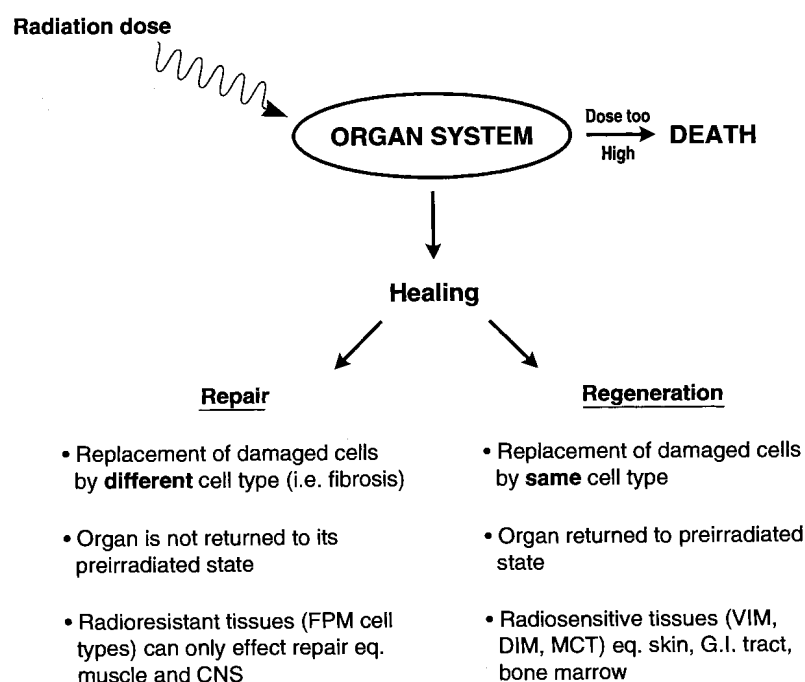


FIGURE 25-11. Schematic diagram of organ system response to radiation.

which case the functionality of the organ system is compromised. The types of response and the degree to which they occur are functions of the dose and the relative radiosensitivity and regenerative capacity of the cells that comprise the organ system. If the exposures are excessive, the ability of the cells to effect any type of healing may be lost, resulting in tissue necrosis.

Characterization of Radiosensitivity

The characterization of an organ system as radioresistant or radiosensitive depends in large part on the radiosensitivity of cells that comprise the functional parenchyma. The cells of the supportive stromal tissue consists mainly of cells of intermediate radiosensitivity. Therefore, when the parenchyma consists of radiosensitive cell types (e.g., stem cells), the initial hypoplasia and concomitant decrease in functional integrity will be the result of damage to these radiosensitive cell populations. Functional changes are typically apparent within days or weeks after the exposure. However, if the parenchyma is populated by radioresistant cell types (e.g., nerve or muscle cells), damage to the functional layer occurs indirectly by compromise of the cells in the vascular stroma. In this case, the hypoplasia of the parenchymal cells is typically delayed several months and is secondary to changes in the supportive vasculature. The principal effect is a gradual proliferation of intimal cells in the microvasculature that eventually narrows and reduces the blood supply to the point that the flow of oxygen and nutrients is insufficient to sustain the functional parenchyma.

Specific Organ System Responses

This section focuses on radiation-induced changes to the skin, reproductive organs, and eyes. Effects on the hematopoietic, gastrointestinal, cardiovascular, and central nervous systems are addressed later.

Skin

The first evidence of biologic effects of ionizing radiation appeared on exposed skin in the form of erythema and acute radiation dermatitis. In fact, before the introduction of the roentgen as the unit of radiation exposure, radiation therapists evaluated the intensity of x-rays by using a quantity called the “skin erythema dose” which was defined as the dose of x-rays necessary to cause a certain degree of erythema within a specified time. This quantity was unsatisfactory for a variety of reasons, not least of which was that the response to skin from radiation exposure is quite variable. The reaction of skin to ionizing radiation has been studied extensively, and the degree of damage has been found to depend on not only the radiation quantity, quality, and dose rate but also on the location and extent of the exposure. With the exception of prolonged, fluoroscopically guided interventional procedures (see Chapter 23), it is highly unlikely that doses from diagnostic examinations will be high enough to produce any of the effects to be discussed in this section.

The most sensitive structures in the skin include the germinal epithelium, sebaceous glands, and hair follicles. The cells that make up the germinal epithelium, located between the dermis and the epidermis, have a high mitotic rate and continuously replace sloughed epidermal cells. Complete turnover of the epidermis normally occurs within approximately 2 weeks. Skin reactions to radiation exposure are deterministic and have a threshold of approximately 1 Gy (100 rad), below which no effects are seen.

Temporary hair loss (epilation) can occur in approximately 3 weeks after exposure to 3 to 6 Gy (300 to 600 rad), with regrowth beginning approximately 2 months later and complete within 6 to 12 months. Erythema occurs within 1 to 2 days after an acute dose of low-LET radiation between 2 and 6 Gy (200 and 600 rad) and lasts for a few days. Higher doses produce earlier and more intense erythema. This type of skin reaction is largely caused by capillary dilatation secondary to the radiation-induced release of vasoactive amines (e.g., histamine). Erythema can reappear several weeks after the initial exposure, reaching a maximal response on about the third week, at which time the skin is edematous, tender, and often exhibits a burning sensation. Recovery is typically complete within 4 to 8 weeks after the exposure. Erythema may be followed by a dry desquamation for acute exposures of approximately 15 Gy (1,500 rad) or 30 Gy (3,000 rad) if the dose is fractionated over several weeks.

After moderately large doses—40 Gy (4,000 rad) over a period of 4 weeks or 20 Gy (2,000 rad) in a single dose—intense erythema followed by an acute radiation dermatitis and moist desquamation occurs and is characterized by edema, dermal hypoplasia, inflammatory cell infiltration, damage to vascular structures, and permanent hair loss.

Provided the vasculature and germinal epithelium of the skin have not been too severely damaged, re-epithelialization occurs within 6 to 8 weeks, returning the skin

to normal within 2 to 3 months. If these structures have been damaged but not destroyed, healing may occur although the skin may be atrophic, hyperpigmented, and easily damaged by minor physical trauma. Recurring lesions and infections at the site of irradiation are common in these cases, and necrotic ulceration can develop. A chronic radiation dermatitis can also be produced by repeated low-level exposures (10 to 20 mGy/day [1 to 2 rad/day]) where the total dose approaches 20 Gy (2,000 rad) or more. In these cases the skin may become hypertrophic or atrophic and is at increased risk for development of skin neoplasms (especially squamous cell carcinoma). Erythema will not result from chronic exposures in which the total dose is less than 6 Gy (600 rad). Table 25-3 summarizes the effects of radiation on skin.

TABLE 25-3. SUMMARY OF RADIATION EFFECTS ON SKIN

Spectrum of effects on skin			
Early effects	Late effects	Effect on accessory structures	
Erythema Inflammation Dry desquamation Moist desquamation	Atrophy Fibrosis Hyper/hypopigmentation Ulceration Necrosis Cancer	Epilation Destruction of sweat and sebaceous glands	
Dose/time—response relationship			
Dose	Radiation dose area	Exposure interval	Type of reaction or damage
<1 Gy (100 rad)	Small	Short	No visible effect
2–6 Gy (200–600 rad)	Small	Short	Erythema in 1–2 days after exposure; persists until day 5–6; reappears on day 10–12; maximum on day 18–20; persists until day 30–40. Temporary hair loss if >3 Gy (300 rad)
6–10 Gy (600–1,000 rad)	Small	Short	More serious erythema caused by damage of basal cells. Symptoms appear earlier, are more intense, and healing is delayed.
15 Gy (1,500 rad) or 30 Gy (3,000 rad)	Small	Short	Severe erythema, followed by dry desquamation and delayed/incomplete healing.
20–50 Gy (2,000–5,000 rad) or 40 Gy (4,000 rad)	Limited	Short	Intense erythema, acute radiation dermatitis with moist desquamation, edema, dermal hypoplasia, vascular damage and permanent hair loss, permanent tanning, destruction of sweat glands, vascular damage. If >5 Gy (5,000 rad) followed by ulceration/necrosis
20 Gy (2,000 rad)	Hands or other small area	Several years small daily doses of 10–20mGy (1–2 rad)	No early or intermediate changes. Late changes manifested by dry cracked skin, nails curled and cracked, intractable ulcers, possible cancerous changes.

Reproductive Organs

In general, the gonads are very radiosensitive. The testes contain both radiosensitive (e.g., spermatogonia) and radioresistant (e.g., mature spermatozoa) cell populations. The primary effect of radiation on the male reproductive system is temporary or permanent sterility after acute doses of approximately 2.5 Gy (250 rad) or 5 Gy (500 rad), respectively. Temporary sterility has been reported after doses as low as 150 mGy (15 rad). Reduced fertility due to decreased sperm count and motility can also be seen after chronic exposures of 20 to 50 mGy/wk (2 to 5 rad/wk) when the total dose exceeds 2.5 Gy (250 rad). These effects are not of concern with diagnostic examinations, because doses exceeding 100 mGy (10 rad) are extremely unlikely.

The ova within ovarian follicles (classified according to their size as small, intermediate, or large) are sensitive to radiation. The intermediate follicles are the most radiosensitive, followed by the large (mature) follicles and the small follicles, which are the most radioresistant. Therefore, after a radiation dose as low as 1.5 Gy (150 rad), fertility may be temporarily preserved owing to the relative radioresistance of the mature follicles, and this may be followed by a period of reduced fertility. Fertility will recur provided the exposure was not high enough to destroy the relatively radioresistant small primordial follicles. Doses in excess of 6 Gy (600 rad) are typically required to produce permanent sterility; however, sterility has been reported after doses as low as 3.2 Gy (320 rad). In either case it seems that higher doses are required to produce sterility in younger women. Another major concern is the induction of genetic mutations and their effect on future generations. This subject is addressed later in the chapter.

Ocular Effects

The lens of the eye contains a population of radiosensitive cells that can be damaged or destroyed by radiation. Insofar as there is no removal system for these damaged cells, they can accumulate to the point at which they cause cataracts. The magnitude of the opacity as well as the probability of its occurrence is proportional to the dose. The latent period is inversely related to dose and typically requires at least 1 year after the exposure. High-LET radiation is more efficient for cataractogenesis by a factor of 2 or more. Acute doses as low as 2 Gy (200 rad) have been shown to produce cataracts in a small percentage of people exposed, whereas doses greater than 7 Gy (700 rad) always produce cataracts. Chronic exposure reduces the efficiency of cataract formation. For example, the threshold for cataract formation when the exposure is protracted over 2 months is 4 Gy (400 rad), and for 4 months it is 5.5 Gy (550 rad). Cataracts among early radiation workers were common because of the extremely high doses resulting from long and frequent exposures from poorly shielded x-ray equipment. Today, radiation-induced cataracts are rare and even the highest exposures to the eyes of radiation workers in a medical setting (typically from fluoroscopic procedures) do not approach the threshold for effects over an occupational lifetime. A unique aspect of cataract formation is that, unlike senile cataracts, which typically develop in the anterior pole of the lens, radiation-induced cataracts begin as a small opacity in the posterior pole and migrate anteriorly.

25.4 ACUTE RADIATION SYNDROME

As previously discussed, the body consists of cells of differing radiosensitivities and a large radiation dose delivered acutely yields greater cellular damage than the same

dose delivered over a protracted period. When the whole body is subjected to a large acute radiation exposure, there is a characteristic clinical response known as the *acute radiation syndrome* (ARS). ARS is an organismal response quite distinct from isolated local radiation injuries such as epilation or skin ulcerations.

The ARS is actually a combination of subsyndromes occurring in stages over a period of hours to weeks after the exposure, as the injury to various tissues and organ systems is expressed. These subsyndromes result from the differing radiosensitivities of these organ systems. In order of their occurrence with increasing radiation dose, the ARS is divided into the hematopoietic, gastrointestinal, and neurovascular syndromes. These syndromes identify the organ system, the damage to which is primarily responsible for the clinical manifestation of disease. The ARS occurs only when the exposure (a) is acute, (b) involves the whole body (or at least a large portion of it), and (c) is from external penetrating radiation, such as x-rays, gamma rays, or neutrons (internal or external contamination with radioactive material will not usually produce the ARS due to the relatively low dose rates).

Sequence of Events

The clinical manifestation of each of the subsyndromes occurs in a predictable sequence of events that includes the *prodromal*, *latent*, *manifest illness*, and, if the dose is not fatal, *recovery* stages (Fig. 25-12). The prodromal symptoms typically begin within 6 hours after the exposure. As the whole-body exposure increases above a threshold of approximately 0.5 to 1 Gy (50 to 100 rad) the prodromal symptoms, which include anorexia, nausea, vomiting, and diarrhea, begin earlier and are more intense. Table 25-4 provides an estimate of doses that would be likely to produce these symptoms in 50% of persons exposed. The time of onset and the severity of these symptoms were used during the initial phases of the medical response to the Chernobyl (Ukraine) nuclear reactor accident to triage patients with respect to their radiation exposures. These symptoms subside during the latent period, whose duration is shorter for higher doses and may last for up to 4 weeks for modest exposures less than 1 Gy (100 rad). The latent period can be thought of as an “incubation period” during which the organ system damage is progressing.

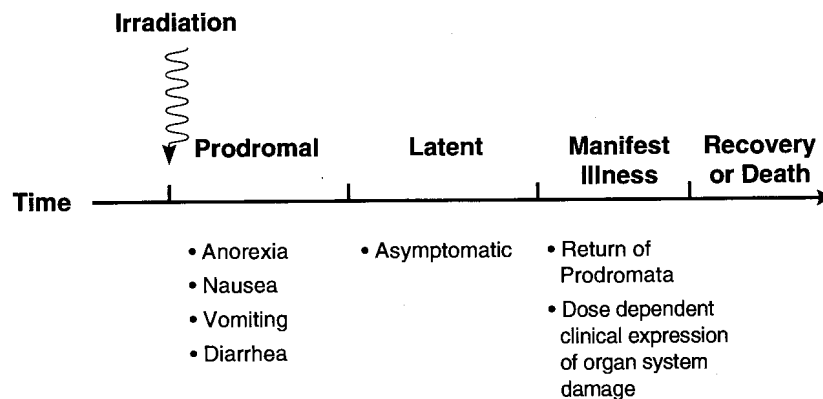


FIGURE 25-12. Phases of acute radiation syndrome.

TABLE 25-4. APPROXIMATE DOSE (IN AIR) THAT WILL PRODUCE VARIOUS PRODROMAL SYMPTOMS IN 50% OF PERSONS EXPOSED

Dose in air	Symptoms
1.2 Gy (120 rad)	Anorexia
1.7 Gy (170 rad)	Nausea
2.1 Gy (210 rad)	Vomiting
2.4 Gy (240 rad)	Diarrhea

Source: Pizzarello DJ, Witcofski RL. *Medical radiation biology*, ed. 2. Philadelphia: Lea & Febiger, 1982.

The latent period ends with the onset of the clinical expression of organ system damage, called the *manifest illness stage*, which can last from 2 to 3 weeks. This stage is the most difficult to manage from a therapeutic standpoint, because of the overlying immunoincompetence that results from damage to the hematopoietic system. Therefore, treatment during the first 6 to 8 weeks after the exposure is essential to optimize the chances for recovery. If the patient survives the manifest illness stage, recovery is almost ensured; however, the patient will be at higher risk for cancer and his or her future progeny for genetic abnormalities.

Hematopoietic Syndrome

The most radiosensitive organ system is the bone marrow, which contains the hematopoietic stem cells. However, with the exception of the lymphocytes, their mature counterparts in circulation are relatively radioresistant.

Hematopoietic tissues are located at various anatomic sites throughout the body; however, posterior radiation exposure maximizes damage because the majority of the active bone marrow is located in the spine and posterior region of the ribs and pelvis. The hematopoietic syndrome is the primary acute clinical consequence of an acute radiation dose between 0.5 to 10 Gy (50 to 1,000 rad). Healthy adults with proper medical care almost always recover from doses lower than 2 Gy (200 rad), whereas doses greater than 8 Gy (800 rad) are almost always fatal unless drastic therapy such as bone marrow transplantation is successful. During the last decade, growth factors such as granulocyte-macrophage colony-stimulating factor and some interleukins have shown promise in the treatment of severe stem cell depletion. Even with effective stem cell replacement therapy, however, it is unlikely that patients will survive doses in excess of 12 Gy (1,200 rad) because of the damage to the gastrointestinal tract and the vasculature. In any case, the human LD_{50/60} (the dose that would be expected to kill 50% of an exposed population within 60 days) is not known with any certainty. The best estimates from the limited data on human exposures are between 3 to 4 Gy (300 to 400 rad) to the bone marrow.

The probability of recovering from a large radiation dose is reduced in patients who are compromised by trauma or other serious concurrent illness. The severe burns and trauma received by some of the workers exposed during the Chernobyl nuclear accident resulted in a lower LD_{50/60} than would have been predicted from their ionizing radiation exposures alone. In addition, patients with certain inherited

diseases such as ataxia telangiectasia, Fanconi's anemia, and Bloom's syndrome are known to have an increased sensitivity to radiation exposure.

The prodromal symptoms occur within a few hours after exposure and may consist of nausea, vomiting, and diarrhea. If these symptoms appear early and severe diarrhea occurs within the first 2 days, the radiation exposure may prove to be fatal. The prodromal and latent periods may each last as long as 3 weeks. Although the nausea and vomiting may subside during the latent period, patients may still feel fatigued and weak. During this period, damage to the stem cells reduces their number and thus their ability to maintain normal hematologic profiles by replacing the circulating blood cells that eventually die by senescence. The kinetics of this generalized pancytopenia are accelerated with higher exposures. Figure 25-13 illustrates the typical hematologic response after bone marrow doses of 1 and 3 Gy (100 and 300 rad). The initial rise in the neutrophil count is presumably a stress response in which neutrophils are released from extravascular stores. The decline in the lymphocyte count occurs within hours after exposure and is a crude early biologic marker of the magnitude of exposure. The threshold

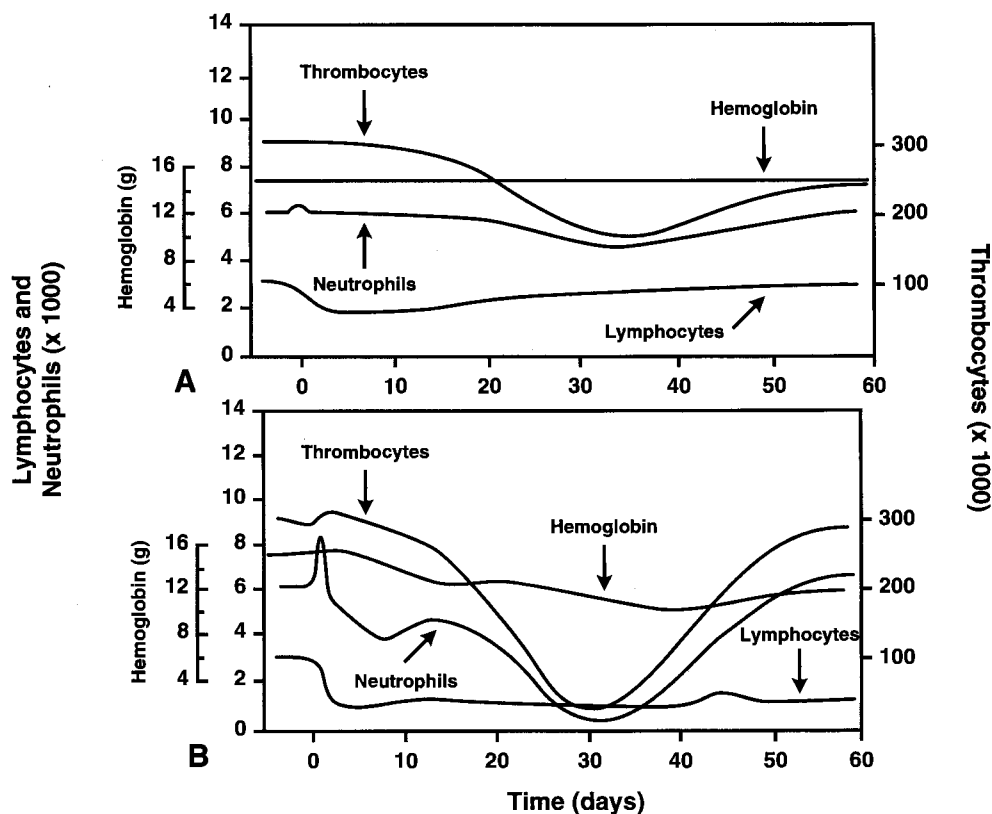


FIGURE 25-13. Typical hematologic depression after bone marrow doses of (A) 1 Gy (100 rad) and (B) 3 Gy (300 rad). (Adapted from Andrews G. Medical management of accidental total-body irradiation. In: Hubner KF, Fry SA, eds. *The medical basis for radiation accident preparedness*. New York: Elsevier, 1980.)

for a measurable depression in the blood lymphocyte count is approximately 0.25 Gy (25 rad); an absolute lymphocyte count lower than $1,000/\text{mm}^3$ in the first 48 hours indicates a severe exposure.

The clinical manifestation of bone marrow depletion occurs 3 to 4 weeks after the exposure as the number of cells in circulation reaches its nadir. Hemorrhage from platelet loss and infection caused by the depletion of white cells are the lethal consequences of severe hematopoietic compromise. Overall, the systemic effects that can occur from the hematopoietic syndrome include mild to profound immunologic compromise, infection, hemorrhage, anemia, and impaired wound healing.

Gastrointestinal Syndrome

At higher exposures the clinical expression of the gastrointestinal syndrome becomes evident, the consequences of which are more immediate, are more severe, and overlap those of the hematopoietic syndrome. At doses greater than 12 Gy (1,200 rad), this syndrome is primarily responsible for lethality. Its prodromal stage includes severe nausea, vomiting, watery diarrhea, and cramps occurring within hours after the exposure, followed by a much shorter latent period (5 to 7 days). The manifest illness stage begins with the return of the prodromal symptoms, often more intensely than in their original presentation. The intestinal dysfunction is the result of the severe damage to the intestinal mucosa. Severely damaged crypt stem cells lose their reproductive capacity. As the mucosal lining ages and eventually sloughs, the cells are not replaced, resulting in a degeneration of the functional integrity of the gut. The result is the loss of capacity to regulate the absorption of electrolytes and nutrients, and at the same time a portal is created for intestinal flora to enter the systemic circulation.

Overall, intestinal pathology includes mucosal ulceration and hemorrhage, disruption of normal absorption and secretion, alteration of enteric flora, depletion of gut lymphoid tissue, and disturbance of gut motility. Systemic effects may include malnutrition resulting from malabsorption; vomiting and abdominal distention from paralytic ileus; anemia from gastrointestinal bleeding; and sepsis from the damage to the structural integrity of the intestinal lining.

The effects of the drastic changes in the gut are compounded by equally drastic changes in the bone marrow. The most significant effect is the severe decrease in circulating white cells at a time when bacteria are invading the bloodstream from the gastrointestinal tract. The other blood cells may not exhibit a significant decrease because death occurs before radiation damage appears in these cell lines. Lethality from the gastrointestinal syndrome is essentially 100%. Death occurs within 3 to 10 days after the exposure if no medical care is given or as long as 2 weeks afterward with intensive medical support.

Neurovascular Syndrome

Death occurs within 2 to 3 days after supralethal doses in excess of 50 Gy (5,000 rad). Doses in this range result in cardiovascular shock with a massive loss of serum and electrolytes into extravascular tissues. The ensuing circulatory problems of edema, increased intracranial pressure, and cerebral anoxia cause death before any other changes in the body become significant.

The stages of the neurovascular syndrome are extremely compressed. The prodromal period may include a burning sensation of the skin that occurs within minutes, followed by nausea, vomiting, confusion, ataxia, and disorientation within 1 hour. There is an abbreviated latent period (4 to 6 hours), during which some improvement is noted, followed by a severe manifest illness stage. The prodromal symptoms return with even greater severity, coupled with respiratory distress and gross neurologic changes (including tremors and convulsions) that inevitably lead to coma and death. Details of some of the other aspects of this syndrome are not well known because human exposures to supralethal radiation are rare. Experimental evidence suggests that the initial hypotension may be caused by a massive release of histamine from mast cells, and the principal pathology may result from massive damage to the microcirculation.

Summary of Acute Radiation Syndrome Interrelationships

Figure 25-14 summarizes the elements and interrelationships of the ARS. Table 25-5 summarizes the ARS with respect to its major components, the dose ranges at which they become significant, the probability of survival, and the principal clinical symptoms.

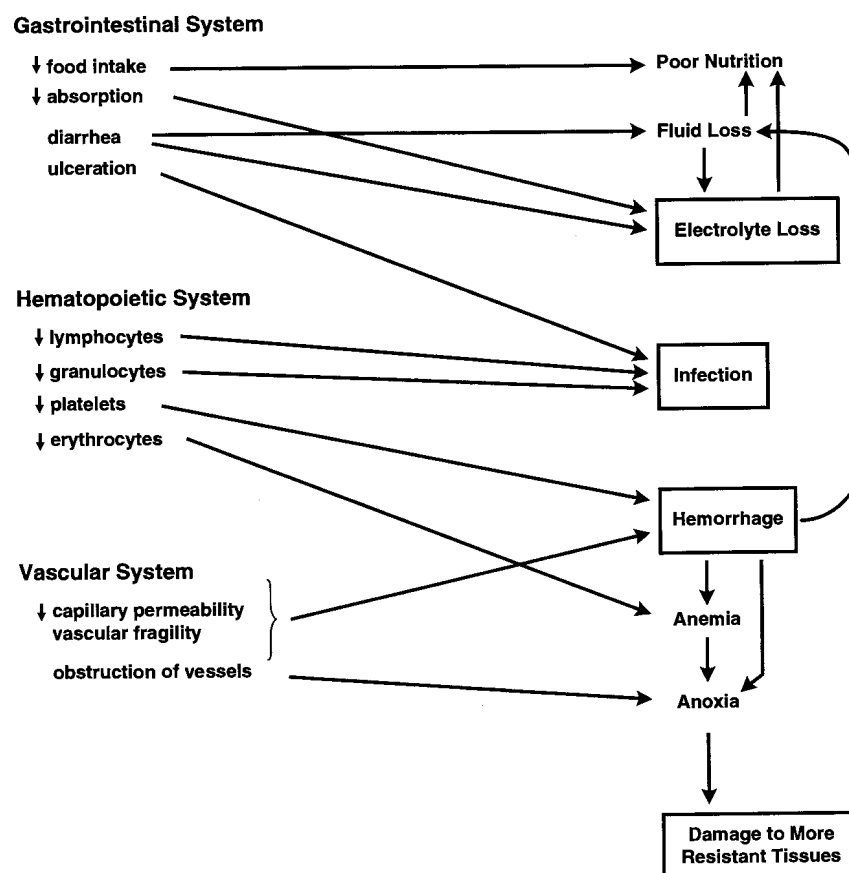


FIGURE 25-14. Interrelationships among various elements of the acute radiation syndrome.

TABLE 25-5. SUMMARY OF THE STAGES, DOSE RANGE, MAJOR CLINICAL SYMPTOMS, AND EFFECT OF MEDICAL INTERVENTION ON THE ACUTE RADIATION SYNDROME

Syndrome	Dose range	Prodromal effects	Manifest-illness effects	Survival without treatment	Survival with treatment
<i>Hematological</i>	0.5–1.0 Gy (50–100 rad)	Mild	Slight decrease in blood cell count	Almost certain	Almost certain
	1.0–2.0 Gy (100–200 rad)	Mild to moderate	Early symptoms of bone-marrow damage	Probable (>90%)	Almost certain
	2.0–3.5 Gy (200–350 rad)	Moderate	Moderate to severe bone-marrow damage	Possible	Probable
	3.5–5.5 Gy ^a (350–550 rad)	Moderate to severe	Severe bone-marrow damage; slight intestinal damage	Death likely within 3–6 wk	Possible
	5.5–8.0 Gy (550–800 rad)	Severe	Bone-marrow pancytopenia and moderate intestinal damage	Death likely within 2–3 wk	Possible if stem cell support or replacement is successful
<i>Gastrointestinal</i>	8.0–10.0 Gy (800–1000 rad)	Severe	Combined gastrointestinal and bone-marrow damage; hypotension	Death within 1.0–2.5 wk	Death likely, but survival possible if stem cell support or replacement is successful
	10.0–50.0 Gy (1,000–5,000 rad)	Severe prodromal syndrome which overlaps with severe gastrointestinal damage; electrolyte imbalance; fatigue, gastrointestinal death		Death within 5–12 days	
<i>Neurovascular</i>	>50 Gy (5,000 rad)	Severe prodromal symptoms which overlap with central nervous system effects, Ataxia, severe edema, neurovascular damage, shock		Death within 2–5 days	

^aThe LD_{50/60} is 3–4 Gy (300–400 rad).

25.5 RADIATION-INDUCED CARCINOGENESIS

Most of the radiation-induced biologic effects discussed thus far are detectable within a relatively short time after the exposure. Ionizing radiation can, however, cause damage whose expression is delayed for years or decades. The ability of ionizing radiation to increase the risk of cancer years after exposure is one example that has been well established. Cancer, however, is not a rare disease; indeed, it is the second most likely cause of death, after cardiovascular disease, in the United States. Cancer incidence and mortality annual rates, averaged for the U.S. population, are approximately 398 and 170 per 100,000, respectively, with males having a higher mortality rate (213 per 100,000) than females (141 per 100,000). Another way of expressing cancer risk is the probability of developing or of dying from cancer, which for the U.S. population is 41% and 22%, respectively (males and females combined). Although the etiologies of cancers are not well defined, diet, lifestyle, and environmental conditions appear to be among the most important factors. For example, the total cancer incidence among populations around the world varies by only a factor of 2 or so, but the incidences of specific cancers can vary by a factor of 200 or more!

Cancer induction is the most important delayed somatic effect of radiation exposure. However, radiation is a relatively weak carcinogen at low doses (e.g., occupational and diagnostic exposures). In fact, the body's capacity to repair some radiation damage means that the possibility of no increased risk at low doses cannot be ruled out. The determinants of radiation-induced cancer risk are discussed in greater detail later in the chapter.

Molecular Biology and Cancer

Cancer arises from abnormal cell division. Cells in a tumor are believed to descend from a common ancestral cell that at some point (typically decades before a tumor results in clinically noticeable symptoms) loses its control over normal reproduction. The malignant transformation of such a cell comes about through the accumulation of mutations in specific classes of genes. Mutations in these genes are a critical step in the development of cancer.

Any protein involved in the control of cell division may also be involved in cancer. However, two classes of genes, *tumor suppressor genes* and *proto-oncogenes*, which respectively inhibit and encourage cell growth, play major roles in triggering cancer. Tumor suppressor genes such as the p53 gene, in its nonmutated or wild-type state, promotes the expression of certain proteins. One of these halts the cell cycle and gives the cell time to repair its DNA before dividing. Alternatively, if the damage cannot be repaired, the protein pushes the cell into a programmed cell death called *apoptosis*. The loss of normal function of the p53 gene product may compromise DNA repair mechanisms and lead to tumor development. Defective p53 genes can cause abnormal cells to proliferate and as many as 50% of all human tumors have been found to contain p53 mutations.

Proto-oncogenes code for proteins that stimulate cell division. Mutated forms of these genes, called *oncogenes*, can cause the stimulatory proteins to be overactive, resulting in excessive cell proliferation. For example, mutations of the Ras proto-oncogenes (H-Ras, N-Ras, K-Ras) are found in about 25% of all human tumors. The Ras family of proteins plays a central role in the regulation of cell growth and

the integration of regulatory signals. These signals govern processes within the cell cycle and regulate cellular proliferation. Most mutations result in abrogation of the normal enzymatic activity of Ras, which causes a prolonged activation state and unregulated stimulation of Ras signaling pathways that either stimulate cell growth or inhibit apoptosis.

Stages of Cancer Development

Cancer is thought to occur as a multistep process in which the initiation of damage in a single cell leads to a preneoplastic stage followed by a sequence of events that permit the cell to successfully proliferate. All neoplasms and their metastases are characterized by unrestrained growth, irregular migration, and genetic diversity and are thought to be derivatives or clones of a single cell. It is assumed that there is no threshold dose for the induction of cancer, because even a single ionization event could theoretically lead to molecular changes in the DNA that result in malignant transformation and ultimately cancer. However, the probability of cancer development is far lower than would be expected from the number of initiating events because of host defense mechanisms and the possibility that all of the subsequent steps required for expression of the malignant potential of the cell will not occur.

Cancer formation can be thought of as occurring in three stages: (a) *initiation*, (b) *promotion*, and (c) *progression*. During initiation, a somatic mutational event occurs that is not repaired. This initial damage can be produced by radiation or any of a variety of other environmental or chemical carcinogens. During the promotion stage, the preneoplastic cell is stimulated to divide. A promoter is an agent that by itself does not cause cancer but, once an initiating carcinogenic event has occurred, promotes or stimulates the cell containing the original damage.

There are more than 60 chemical or physical agents known to be carcinogenic in humans, including benzene, formaldehyde, vinyl chloride, and asbestos. Another 200 or more agents are classified as “reasonably anticipated to be human carcinogens.” Unlike most carcinogens, radiation may act as an initiator or a promoter. Some hormones act as promoters by stimulating the growth of target tissues. For example, estrogen and thyroid-stimulating hormone may act as promoters of breast cancer and thyroid cancer, respectively.

The final stage is progression, during which the transformed cell produces a number of phenotypic clones, not all of which are neoplastic. Eventually, one phenotype acquires the selective advantage of evading the host’s defense mechanisms, thus allowing the development of a tumor and possibly a metastatic cancer. Radiation may also enhance progression by immunosuppression.

Environmental factors implicated in the promotion of cancer include tobacco, alcohol, diet, sexual behavior, air pollution, and bacterial and viral infections. Support for the role of environmental factors comes from observations such as the increased incidence of colon cancer among Japanese immigrants to the United States compared with those living in Japan. Any agents that compromise the immune system, such as the human immunodeficiency virus (HIV), increase the probability of successful progression of a preneoplastic cell into cancer. A number of chemical agents, when given alone, are neither initiators nor promoters but when given in the presence of an initiator will enhance cancer development. Many of these agents are present in cigarette smoke, which may in part account for its potent carcinogenicity. Environmental exposure to nonionizing

radiation, such as radiofrequency radiation from cellular telephones and their base stations or electric and magnetic field exposures from power lines, although it has received considerable media attention, has not yielded reproducible evidence of carcinogenic potential.

Modifiers of Radiation-Induced Cancer

Radiation-induced cancers can occur in most tissues of the body and are indistinguishable from those that arise from other causes. The probability of development of a radiation-induced cancer depends on a number of factors including the dose rate and the radiation quantity and quality. As seen with cells in culture, high-LET radiation is more effective in producing biologic damage than is low-LET radiation. Although the RBE has not been accurately determined, high-LET radiation has been shown to produce more leukemia and cancers of the breast, lung, liver, thyroid, and bone than an equal dose of low-LET radiation.

Fractionation of large doses increases the probability of cellular repair and thereby reduces the incidence of carcinogenesis in some cases such as leukemia. However, at the low doses associated with diagnostic examinations and occupational exposures, dose rate does not appear to affect cancer risk.

Latency and the incidence of some radiation-induced cancers vary with age at the time of exposure. For example, people who are younger than 20 years of age at the time of exposure are at greater risk for development of breast and thyroid cancer but at lower risk for stomach cancer and leukemia, compared with older individuals. The minimal latent period is 2 to 3 years for leukemia, with a period of expression (i.e., the time interval required for full expression of the radiogenic cancer increase) proportional to the age at the time of exposure, ranging from approximately 12 to 25 years. Latent periods for solid tumors range from 5 to 40 years, with a period of expression in some cases longer than 50 years.

Epidemiologic Investigations of Radiation-Induced Cancer

Although the dose-response relationship for cancer induction at high dose (and dose rate) has been fairly well established, the same cannot be said for the low doses resulting from diagnostic and occupational exposures. Insufficient data exist to determine accurately the risks of low-dose radiation exposure to humans. Animal and epidemiologic investigations indicate that the risks of low-level exposure are small, but how small is still a matter of debate in the scientific community.

The populations that form the bases of the epidemiologic investigation of radiation bioeffects come from four principal sources: (a) survivors of the atomic bomb detonations in Hiroshima and Nagasaki; (b) patients with medical exposure during treatment of a variety of neoplastic and nonneoplastic diseases; (c) persons with occupational exposures; and (d) populations with high natural background exposures (Table 25-6). It is very difficult to detect a small increase in the cancer rate due to radiation, because the natural incidence of many forms of cancer is high and the latent period for most cancers is long. To rule out simple statistical fluctuations, a very large irradiated population is required. Some epidemiologic investigations have been complicated by such factors as failure to

TABLE 25-6. SOURCES OF DATA ON RADIATION EXPOSURE TO HUMANS**Atomic-bomb detonation exposures and fallout**

Survivors
 Offspring of survivors

Medical exposures

Treatment of tinea capitis
 X-ray treatment of ankylosing spondylitis
 Prenatal diagnostic x-ray
 X-ray therapy for enlarged thymus glands
 Fluoroscopic guided pneumothorax for the treatment for tuberculosis
 Thorotrast (radioactive contrast material used in angiography, 1925–1945)
 Treatment of neoplastic diseases (e.g., breast cancer, Wilms' tumor, cancer of the cervix and leukemia)

Occupational exposures

Radium dial painters (1920s)
 Uranium miners and millers
 Nuclear dockyard workers
 Nuclear-materials enrichment and processing workers
 Participants in nuclear weapons tests
 Construction workers
 Industrial photography workers
 Radioisotope production workers
 Reactor personnel
 Civil aviation and astronautic personnel
 Phosphate fertilizer industry workers
 Scientific researchers
 Diagnostic and therapeutic radiation medical personnel

Epidemiologic comparisons of areas with high background radiation

control for exposure to other known carcinogens, an inadequate period of observation to allow for full expression of cancers with long latent periods, the use of inappropriate control populations, and (in retrospective studies) incomplete or inaccurate health records or the use of questionnaires that relied on people's recollections to estimate dosage. Table 25-7 summarizes details of some of the principal epidemiologic investigations on which current dose-response estimates are based.

Risk Estimation Models***Dose-Response Curves***

Within the context of the limitations described, scientists have developed dose-response models to predict the risk of cancer in human populations from exposure to low levels of ionizing radiation. These models lead to dose-response curves whose shape can be characterized as nonthreshold *linear*, *linear-quadratic*, and *quadratic*. Figure 25-15 shows two of these dose-response curves (linear and linear-quadratic). A nonthreshold linear curve is more likely to overestimate the incidence of cancer at lower doses from low-LET radiation and therefore is used in radiation protection guidance for estimating the overall cancer risk from occupational and diagnostic exposures. The linear-quadratic dose-response model predicts a lower incidence of cancer than the linear model at low doses and a higher incidence at intermediate

TABLE 25-7. SUMMARY OF MAJOR EPIDEMIOLOGIC INVESTIGATIONS THAT FORM THE BASIS OF CURRENT CANCER DOSE-RESPONSE ESTIMATES IN HUMAN POPULATIONS*

Population and exposure	Effects observed	Strengths and limitations
Atomic-bomb survivors: A mortality study of approximately 120,000 residents of Hiroshima and Nagasaki (1950) among which 93,000 were exposed at the time of the bombing. This group continues to be followed. The mortality assessment through 1987 has been completed. New dose estimates were made which determined that neutrons were not, as previously thought, a significant component of the dose for either of the two cities. Mean organ doses have been calculated for twelve organs. Approximately 20,000 received doses of 10–50 mGy (1–5 rad) while ~1,100 received doses in excess of 2 Gy (200 rad).	A total of 483 excess cancers are thought to have been induced by radiation exposure resulting in an excess cancer mortality of 205. The natural incidence for this population would have predicted 2,670 cancers. The number of expected, observed, and excess cancers by organ system is shown below. These data apply to the 41,000 persons who received 0.01 Sv (1 rem) or more. The mean dose was 0.23 Sv (23 rem). No statistically significant radiation-related increases were observed for uterus, cervix, pancreas, rectum, multiple myeloma, non-Hodgkin's lymphoma, or chronic lymphocytic leukemia.	The analysis of the atom-bomb survivors is the single most important cohort which influence current radiation-induced cancer risk estimates. The population is large and there is a wide range of doses from which it is possible to make determinations of the dose-response and the effects of modifying factors such as age on the induction of cancer. Data at high doses are limited, thus the analysis only included individuals in which the doses to internal organs were less than 4 Gy (400 rad). The survivors were not representative of a normal Japanese population insofar as many of the adult males were away on military service while those remaining presumably had some physical condition preventing them from active service. In addition, children and the elderly perished shortly after the detonation in greater numbers than did young adults, suggesting the possibility that the survivors may represent a harder subset of the population.
Incidence data		
Cancer Type	Expected	Observed Excess
Breast	200	295 95
Lung	365	456 91
Stomach	1222	1307 85
Leukemia	67	141 74
Thyroid	94	132 38
Colon	194	223 29
Skin	76	98 22
Bladder	98	115 17
Liver	95	109 14
Esophagus	79	84 5
Brain and central nervous system	67	71 4
Bone and connective tissue	12	16 4
Non-Hodgkin's Lymphoma	72	76 4
Multiple myeloma	29	30 1
TOTAL	2,670	3,153 483

Ankylosing spondylitis: This cohort consists of approximately 14,000 patients treated with radiotherapy to the spine for ankylosing spondylitis throughout the United Kingdom between 1935 and 1954. Although individual dose records were not available, estimates were made which ranged from 1–25 Gy (100–2,500 rad) to the bone marrow and other various organs.	Mortality has been reported though 1982, at which point 727 cancer deaths had been reported. Excess leukemia rates were reported from which an absolute risk of 80 excess cases/Gy (per 100 rad)/year per million was estimated.	This group represents one of the largest bodies of data on radiation-induced leukemia in humans for which fairly good dose estimates exist. Control groups were suboptimal, however, and doses were largely unfractionated. In addition only cancer mortality (not incidence) was available for this cohort.
Postpartum mastitis study: This group consists of approximately 600 women, mostly between the ages of 20 and 40 years, treated with radiotherapy for postpartum acute mastitis in New York in the 1940s and 1950s for which ~1,200 nonexposed women with mastitis and siblings of both groups of women served as controls. Breast tissue doses ranged from 0.6–14 Gy (60–1,400 rad).	Forty-five year follow-up identified excess breast cancer in this population as compared with the general female population of New York.	A legitimate objection to using the data from this study to establish radiation-induced breast cancer risk factors is the uncertainty as to what effect the inflammatory changes associated with postpartum mastitis and the hormonal changes due to pregnancy have on the risk of breast cancer.
Radium dial painters: Young women who ingested radium (Ra-226 and Ra-228 with half-lives of approximately 1600 and 7 years, respectively) while licking their brushes (containing luminous radium sulfate) to a sharp point during the application of luminous paint on dials and clocks in the 1920s and 1930s. Over 800 followed.	Large increase in osteogenic sarcoma. Osteogenic sarcoma is a rare cancer (incidence, ~5 per 10⁵ population). Relative risk in the population was >100x's. No increase was seen below doses of 5 Gy (500 rad) but rose sharply thereafter.	One of only a few studies that analyzed the radiocarcinogenic effectiveness of internal contamination with high LET radiation in humans.
Thorotrast: Several populations were studied in which individuals were injected intravascularly with an x-ray contrast media, Thorotrast, used between 1925 and 1945. Thorotrast contains 25% by weight radioactive colloidal Th-232 dioxide. Th-232 is an alpha emitter with a half-life of approximately 14 billion years.	Particles were deposited in the reticuloendothelial systems. Noted increase in number of liver cancers, particularly angiosarcoma, bile duct carcinomas, and hepatic cell carcinomas. Evaluation of the data resulted in estimates of alpha radiation-induced liver cancer risk of approximately 300/10⁴ per Gy (100 rad) which appears to be linear with dose.	Dose estimates are fairly good. However, the extent to which the chemical toxicity of the Thorotrast may have influenced the risk is not known.

Source: Adapted from National Academy of Sciences/National Research Council Committee on the Biological Effects of Ionizing Radiation. *The health effects of exposure to low levels of ionizing radiation (BEIR V)*. Washington, DC: NAS/NRC, 1990.

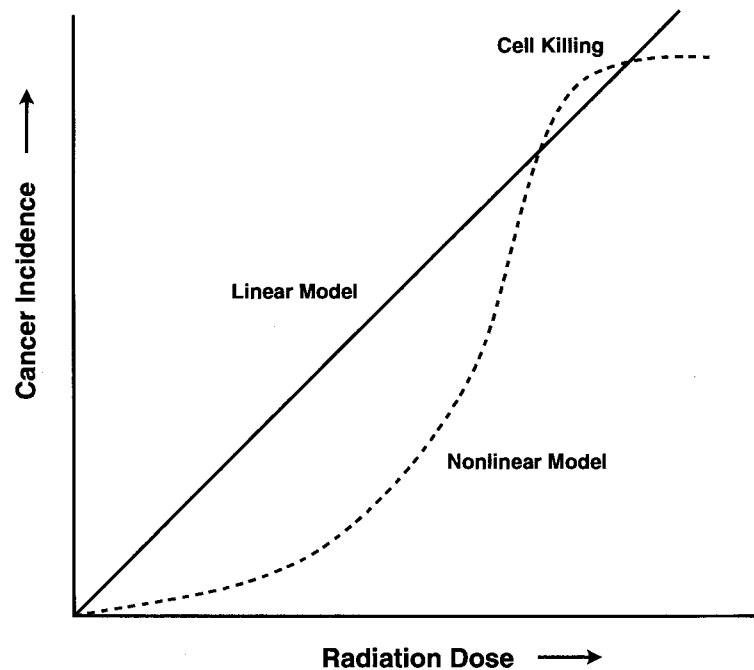


FIGURE 25-15. Linear and (nonlinear-quadratic) dose-response models for the induction of cancer from low linear energy transfer (LET) radiation exposure.

doses. A more general form of this curve predicts a decline in incidence at high doses due to cell killing, which reduces the population of cells available for neoplastic transformation. Risk estimates were revised in 1990 owing principally to a reassessment of the doses received by the Japanese atomic bomb survivors and an additional decade of follow-up of the survivors. The result was an increase in the risk estimates because of (a) revised neutron dosimetry, which implied a greater risk from low-LET radiation, (b) an increase in the number of cancers among the survivors, and (c) use of a revised model that projects risks beyond the period of observation.

Risk Models

Two simple models of the risk of cancer after an exposure are the *multiplicative* and the *additive* risk models. The influence of age at the time of exposure on the cancer risk estimate is accounted for in a multiplicative risk model. The multiplicative risk model predicts that, after the latent period, the excess risk is a multiple of the natural age-specific cancer risk for the population in question (Fig. 25-16A). The projected risk varies with the natural incidence of the cancer in question, with the general trend that *risks increase with age*. Figure 25-17 illustrates the differences in magnitude of the projected radiogenic cancer risk from a given exposure at different ages.

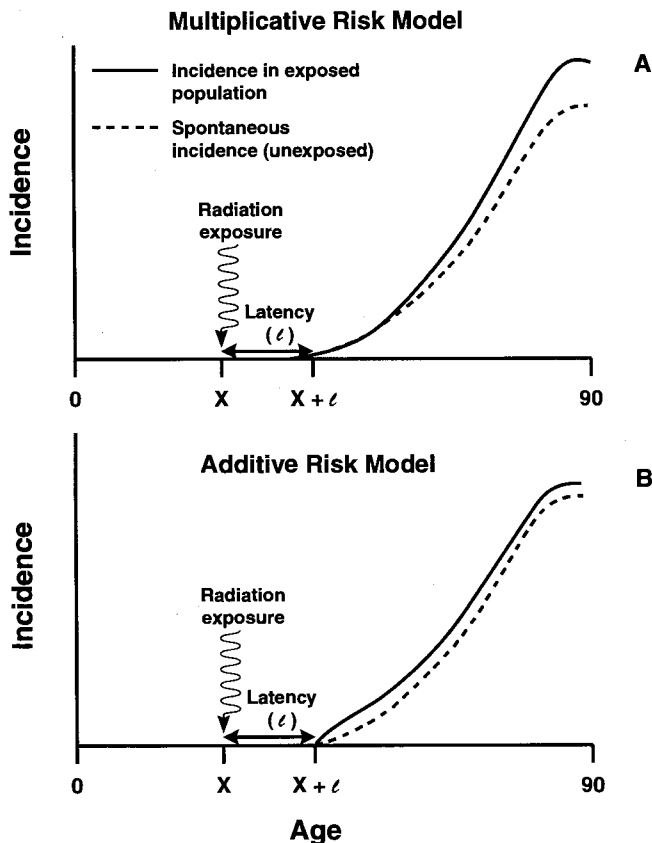


FIGURE 25-16. Comparison of multiplicative and additive risk models. Radiation-induced risk increments are seen after a minimal latent period l ; x is age at exposure. When the increased risk is assumed to occur over the remaining lifetime, induced effects can follow either the multiplicative risk model (**A**), in which the excess risk is a multiple of the natural age-specific cancer risk for a given population, or the additive risk model (**B**), in which a constant increment in incidence is added to the spontaneous disease incidence throughout life. The multiplicative risk model predicts the greatest increment in incidence at older ages.

An alternative hypothesis is the additive risk model, which predicts a fixed (or constant) increase in risk unrelated to the spontaneous age-specific cancer risk at the time of exposure (see Fig. 25-16B).

In general, neither of these two simple models appears to adequately describe the risk of cancer after exposure. A modification of the relative risk model, with corrections for factors such as type of cancer, gender, age at time of exposure, and time elapsed since exposure, was recommended by expert committees on radiation-induced cancer risks (see later discussion of BEIR V risk estimates).

Risk Expression

One way of expressing the risk from radiation in an exposed population is the *relative risk* (RR). Relative risk is the ratio of the cancer incidence in the exposed population to that in the general (unexposed) population; thus, a relative risk of 1.2 would indicate a 20% increase over the spontaneous rate that would otherwise have been expected. The excess relative risk is simply $RR - 1$; in this case, $1.2 - 1 = 0.2$.

To be able to detect a relative cancer risk of 1.2 with a statistical confidence of 95% (i.e., $p < .05$) when the spontaneous incidence is 2% in the population (typi-

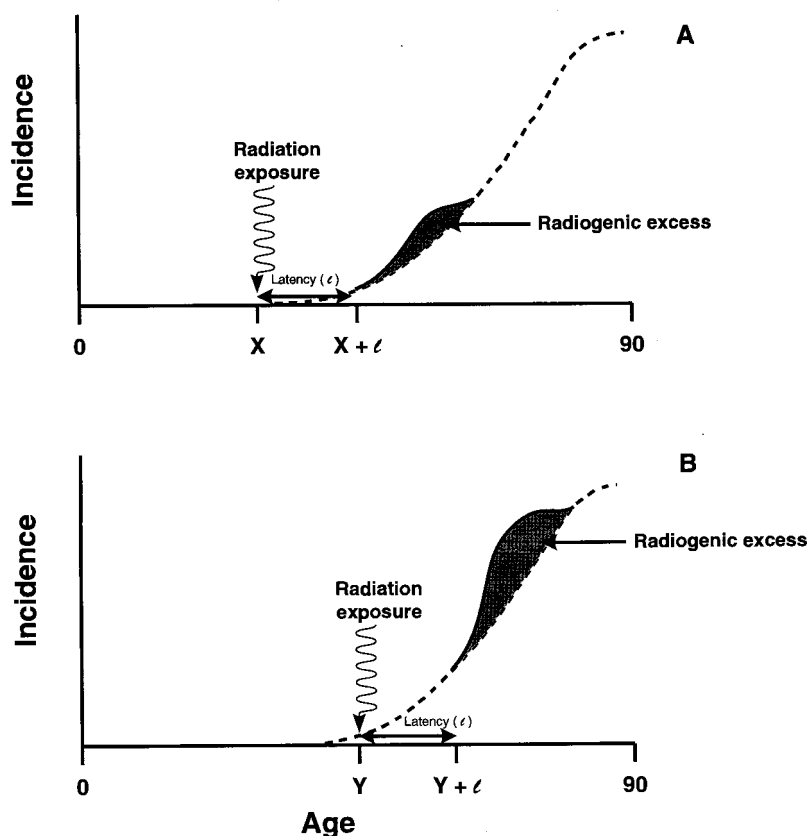


FIGURE 25-17. Multiplicative risk model. The radiation-induced cancer effect is superimposed on spontaneous cancer incidence by age for effects that have a limited expression (e.g., leukemia, bone cancer). Irradiation at younger age (**A**) and at older age (**B**) are shown; x is the age at exposure, and ℓ is the minimal latent period. Predictions made on the basis of the multiplicative risk model typically result in larger increments in cancer incidence when exposure occurs at older ages, where the spontaneous risk of cancer is high and the radiogenic excess is greater than after exposures earlier in life.

cal of many cancers), a study population in excess of 10,000 is required. More than 1 million people would be required to identify a relative risk of 1.01 (i.e., a 1% cancer rate increase) in this same population!

Absolute risk is another way of expressing risk. In this case, risk is expressed as the number of excess radiation-induced cancers per 10^4 people/Sv-yr. For a cancer with a radiation-induced risk of 4 per 10,000 person-Sv and a latency period of about 10 years, the risk of developing cancer from a dose of 0.1 Sv (10 rem) within the next 40 years would be $40 - 10$ or 30 years $\times 0.1$ Sv $\times 4$ per 10,000 person-Sv = 12 per 10,000 or 0.12%. In other words, if 10,000 randomly selected people each received a dose of 0.1 Sv (10 rem), 12 additional cases of cancer would be expected to develop in that population over the subsequent 40 years.

BEIR V Risk Estimates

The National Academy of Sciences/National Research Council Committee on the Biological Effects of Ionizing Radiation (BEIR) published a report in 1990 entitled, *The Health Effects of Exposure to Low Levels of Ionizing Radiation*, commonly referred to as the *BEIR V report*. The BEIR V cancer risk estimate is 8% per Sv (0.08% per rem), which applies to *high doses* delivered at a *high dose rate*. A dose and dose-rate effectiveness reduction factor of approximately 2 was used to correct risk estimates for exposure conditions typically encountered in medical and occupational settings. Therefore, the *single best estimate of radiation-induced mortality at low exposure levels is 4% per Sv (0.04% per rem)*. The risk estimates of the International Commission on Radiological Protection (ICRP), derived from evaluation of the radiation-induced mortality data, are in general agreement with those in the BEIR V report. The ICRP estimates the radiation-induced mortality at low doses to be 4% per Sv (0.04% per rem) for a population of adult workers and 5% per Sv (0.05% per rem) for the whole population (which includes more radiosensitive sub-populations, such as the young).

The BEIR V Committee believed that the linear dose-response model was best for all cancers except leukemia and bone cancer; for those malignancies, a linear-quadratic model was recommended. According to the linear risk-projection model, an exposure of 10,000 people to 10 mSv (1 rem) would result in approximately 4 cancer deaths in addition to the 2,200 (i.e., 22%) normally expected in the general population. This represents an increase of less than 0.2% beyond the spontaneous cancer mortality rate for every 10 mSv (1 rem) of exposure. Table 25-8 lists

TABLE 25-8. OVERALL RISK AND RELATIVE PROBABILITIES OF RADIATION-INDUCED CANCER MORTALITY IN VARIOUS ORGANS FOR DIFFERENT AGE GROUPS^a

Organ	Age group		
	0–19 Yr	0–90 Yr	20–64 Yr
Stomach	0.266	0.291	0.305
Lung	0.192	0.174	0.159
Bone marrow	0.052	0.077	0.109
Bladder	0.030	0.052	0.082
Colon	0.255	0.180	0.089
Esophagus	0.021	0.038	0.061
Ovary ^b	0.009	0.014	0.023
Breast ^b	0.025	0.023	0.022
Remainder	0.150	0.150	0.150
Total	1.000	1.000	1.000
Risk 10 ⁻² /Sv (10 ⁻⁴ /rem)	12.3	5.4	4

^aThis data applies to a Japanese population of a wide range of ages. A multiplicative risk projection model was used. A dose and dose rate effectiveness factor (DDREF) of 2 was applied to the data acquired at high dose and dose rate to adjust for typical occupational and diagnostic exposures, (i.e., per ICRP report 60 recommendation for doses <20 rad or dose rates <10 rad/hr, p. 18, par. 74).

^bThis table represents the ICRP's risk coefficients for a working population that is assumed to be half women and half men. The risks for ovaries and breasts for women alone is therefore twice this risk. No coefficient was specifically given for testis, which is included in the remainder risk coefficient.

Source: Adapted from ICRP Publication 60, *1990 Recommendations of the International Commission on Radiological Protection*. Annals of the ICRP 21 No. 1–3, 1991.

the overall risk and relative probabilities of radiation-induced cancer mortality in various organs for different age groups. Note that at young ages (0 to 19 years) the probability of radiation-induced cancer is a few times higher than the nominal value in the population as a whole. Conversely, the ICRP estimates indicate that the risk decreases after radiation exposure at ages beyond about 50 years, reaching values equal to one-fifth to one-tenth of the nominal value at ages of 70 to 80 years. An additional perspective on cancer risks is presented in Table 25-9, which compares adult cancers with respect to their spontaneous and radiation-induced incidence.

In 1999 the National Academy of Sciences established a committee to revise the BEIR V report; The Committee on Health Risks from Exposure to Low Levels of Ionizing Radiation (BEIR VII). This committee was formed to analyze the large amount of data published since 1990 on the risks to humans of exposure to low-level, low-LET radiation. The committee was to consider relevant data derived from molecular, cellular, animal, and human epidemiologic studies in its comprehensive reassessment of the health risks resulting from exposures to ionizing radiation. The BEIR VII report is due to be published in late 2003.

TABLE 25-9. SPONTANEOUS INCIDENCE AND SENSITIVITY OF VARIOUS TISSUES TO RADIATION-INDUCED CANCER

Site or type of cancer	Spontaneous incidence	Radiation sensitivity
Most Frequent Radiation-Induced Cancers		
Female breast	Very high	High
Thyroid	Low	Very high, especially in females
Lung (bronchus)	Very high	Moderate
Leukemia	Moderate	Very high
Alimentary tract	High	Moderate low
Less Frequent Radiation-Induced Cancers		
Pharynx	Low	Moderate
Liver and biliary tract	Low	Moderate
Lymphomas	Moderate	Moderate
Kidney and bladder	Moderate	Low
Brain and nervous system	Low	Low
Salivary glands	Very low	Low
Bone	Very low	Low
Skin	High	Low
Magnitude of Radiation Risk Uncertain		
Larynx	Moderate	Low
Nasal sinuses	Very low	Low
Parathyroid	Very low	Low
Ovary	Moderate	Low
Connective tissue	Very low	Low
Radiation Risk Not Demonstrated		
Prostate	Very high	Absent?
Uterus and cervix	Very high	Absent?
Testis	Low	Absent?
Mesothelium	Very low	Absent?
Chronic lymphatic leukemia	Low	Absent?

Source: Modified from Committee on Radiological Units, Standards and Protection: *Medical Radiation: A Guide to Good Practice*. Chicago: American College of Radiology, 1985.

Overview of Specific Cancer Risk Estimates

Leukemia

Leukemia is a relatively rare disease, with an incidence in the general U.S. population of approximately 1 in 10,000. However, genetic predisposition to leukemia can dramatically increase the risk. For example, an identical twin sibling of a leukemic child has a 1 in 3 chance of developing leukemia. Although it is rare in the general population, leukemia is one of the most frequently observed radiation-induced cancers, accounting for approximately one-sixth of the mortality from radiocarcinogenesis. Leukemia may be acute or chronic, and it may take a lymphocytic or myeloid form. With the exception of chronic lymphocytic leukemia, increases in all forms of leukemia have been detected in human populations exposed to radiation and in animals experimentally irradiated.

An increase in the incidence of leukemia first appeared among the atomic-bomb survivors 2 to 3 years after the detonation and reached a peak approximately 12 years after the exposure. The incidence of leukemia is influenced by age at the time of exposure (Fig. 25-18). The younger the person is at the time of exposure, the shorter are the latency period and the period of expression. Although the incidence of radiation-induced leukemia decreases with age at the time of exposure, so does the interval of increased risk. The BEIR V Committee recommended a relative-risk, nonlinear dose-response model for leukemia that predicts an excess lifetime risk of 10 in 10,000 (i.e., 0.1%) after an exposure to 0.1 Gy (10 rad).

Thyroid Cancer

Thyroid cancer is also of concern after radiation exposure. It accounts for approximately 6% to 12% of the mortality attributed to radiation-induced cancers. Studies of exposed human populations indicate that females are at approximately 3 to 5

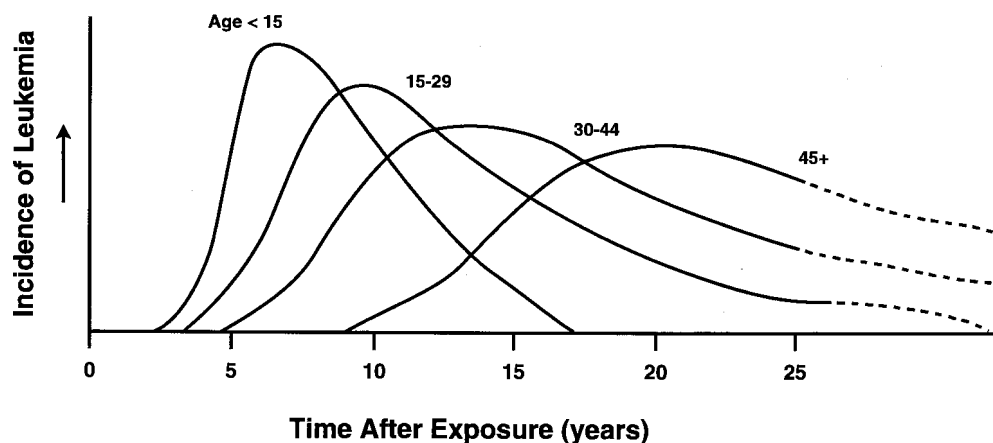


FIGURE 25-18. Effect of age at the time of exposure on the incidence of leukemia (all forms except chronic lymphocytic leukemia) among the atomic-bomb survivors. (From Ichimaru et al. *Incidence of leukemia in atomic-bomb survivors, Hiroshima and Nagasaki 1950–1971, by radiation dose, years after exposure, age, and type of leukemia*. Technical Report RERF 10-76. Hiroshima, Japan: Radiation Effects Research Foundation, 1976.)

times greater risk for development of radiation-induced thyroid cancer than males, presumably because of hormonal influences on thyroid function. Persons of Jewish and North African ancestry also appear to be at greater risk than the general population. The majority of radiation-induced thyroid neoplasms are well-differentiated papillary adenocarcinomas, with a lower percentage being of the follicular form. Because radiation-induced thyroid cancers do not usually include the anaplastic and medullary types, the associated mortality rate is only approximately 5%.

The latency period for benign nodules is 5 to 35 years, and for thyroid malignancies it is 10 to 35 years. The dose-response data for thyroid cancer fit a linear pattern. Studies comparing the sources of radiation exposure indicate that, in general, internal irradiation from radioactive material such as iodine-131 is substantially less effective in producing cancer than the same dose delivered via external irradiation.

Irradiation of the thyroid may produce other responses, such as hypothyroidism and thyroiditis. Threshold estimates for adult populations range from 2 Gy (200 rad) for external irradiation to 50 Gy (5,000 rad) for internal (low-dose-rate) irradiation. Lower threshold estimates exist for children. Approximately 10% of persons with internal thyroid doses of 200 to 300 Gy (20,000 to 30,000 rad) to the thyroid gland from radioactive iodine will develop symptoms of thyroiditis and/or a sore throat, and higher doses may result in thyroid ablation.

The Chernobyl nuclear plant accident released large quantities of radioiodine which resulted in a substantial increase in the incidence of thyroid cancer among children living in heavily contaminated regions. The increase so far has been almost entirely papillary carcinoma, and the incidence rate falls to normal levels among children born more than 6 months after the accident. So far, no other cancers have shown an increased incidence, and there was no identifiable increase in the incidence of birth defects.

Breast Cancer

For women exposed to low-level ionizing radiation, breast cancer is of major concern because of its high radiogenic incidence and high mortality rate. One out of every 8 women in the United States is at risk of developing breast cancer (approximately 180,000 new cases of breast cancer occur each year in the United States), and approximately 1 out of every 30 women is at risk of dying from breast cancer. Some of the etiologic factors in the risk of breast cancer include age at first full-term pregnancy, family history of breast cancer, race, and estrogen levels in the blood. Women who have had no children or only one child are at greater risk for breast cancer than women who have had two or more children. In addition, reduced risks are associated with women who conceive earlier in life and who breast-feed for a longer period of time. Familial history of breast cancer can increase the risk twofold to fourfold, with the magnitude of the risk increasing as the age at diagnosis in the family member decreases. There is also considerable racial and geographic variation in breast cancer risk. Several investigations have suggested that the presence of estrogen acting as a promoter is an important factor in the incidence and latency associated with spontaneous and radiation-induced breast cancer.

The breast cancer risk for low-LET radiation appears to be age dependent, being approximately 50 times higher in the 15-year-old age group (approximately 30/10,000 or 0.3% per year) after an exposure of 0.1 Gy (10 rad) than in women

older than 55 years of age. The risk estimates for women in the 25-, 35-, and 45-year-old age groups are 0.05%, 0.04%, to 0.02%, respectively, according to the BEIR V report. The data fit a linear dose-response model, with a dose of approximately 0.8 Gy (80 rad) required to double the natural incidence of breast cancer. The data from acute and chronic exposure studies indicate that fractionation of the dose reduces the risk of breast cancer induced by low-LET radiation. The latent period ranges from 10 to 40 years, with younger women having longer latencies. In contrast to leukemia, there is no identifiable window of expression; therefore, the risk seems to continue throughout the life of the exposed individual.

Improvements in mammography have resulted in a substantial reduction in the dose to the breast. Women participating in large, controlled mammographic screening studies have been shown to have a decreased risk of mortality from breast cancer. The American Cancer Society and the American College of Radiology currently recommend annual mammography examination for women beginning at age 40 years.

25.6 HEREDITARY EFFECTS OF RADIATION EXPOSURE

Conclusive evidence of the ability of ionization radiation to produce genetic effects was first obtained in 1927 with the experimental observations of radiation-induced genetic effects in fruit flies. Extensive laboratory investigations since that time (primarily large-scale studies in mice) have led scientists to conclude that (a) radiation is a potent mutagenic agent; (b) most mutations are harmful to the organism; (c) radiation does not produce unique mutations; and (d) radiation-induced genetic damage can theoretically occur (like cancer) from a single mutation and appears to be linearly related to dose (i.e., linear nonthreshold dose-response model). Although genetic effects were initially thought to be the most significant biologic effect of ionizing radiation, it is clear that, for doses associated with occupational and medical exposure, the risks are small compared with the spontaneous incidence of genetic anomalies and are secondary to their carcinogenic potential.

Radiation-Induced Genetic Effects in Humans

Epidemiologic investigations have failed to demonstrate radiation-induced genetic effects, although mutations of human cells in culture have been shown. For a given exposure, the mutation rates found in the progeny of irradiated humans are significantly lower than those previously identified in insect populations. The largest population studied is the atomic-bomb survivors and their progeny. Based on current risk estimates, failure to detect an increase in radiation-induced mutations in this population is not surprising considering how few are predicted in comparison to the spontaneous incidence. Screening of 28 specific protein loci in the blood of 27,000 children of atomic-bomb survivors resulted in only two mutations that might have been caused by radiation exposure of the parents.

Earlier studies of survivors' children to determine whether radiation exposure caused an increase in sex-linked lethal genes that would have resulted in increased prenatal death of males or alteration of the gender birth ratio were negative. Irradiation of human testes has been shown to produce an increase in the incidence of translocations, although no additional chromosomal aberrations have been detected in children of atomic-bomb survivors.

Estimating Genetic Risk

The genetically significant dose (GSD) is an index of the presumed genetic impact of radiation-induced mutation in germ cells in an exposed population (see GSD definition in Chapter 23). The sensitivity of a population to radiation-induced genetic damage can be measured by the *doubling dose*, defined as the dose required per generation to double the spontaneous mutation rate. The spontaneous mutation rate is approximately 5×10^{-6} per locus and 7 to 15×10^{-4} per gamete for chromosomal abnormalities. The doubling dose for humans is estimated to be at least 1 Gy (100 rad) per generation; however, this represents an extrapolation from animal data.

The BEIR V Committee estimated that an exposure of 10 mGy (1 rad) to the present generation would cause 6 to 65 additional genetic disorders per 1 million

TABLE 25-10. ESTIMATES OF EXCESS GENETIC EFFECTS AFTER DOSES OF 10 mSv (1 rem) PER GENERATION^a

Type of disorder	Current incidence per 10 ⁶ Live-born offspring	Additional cases per 10 ⁶ Live-born offspring per 10 mSv (1 rem) per generation	
		First generation	Equilibrium
Autosomal dominant ^b			
Clinically severe	2,500	5–20	25
Clinically mild	7,500 ^c	1–15	75
X-linked	400	<1	<5
Recessive	2,500	<1	VSI
Chromosomal			
Unbalanced translocations	600 ^d	<5	VLI
Trisomies	3,800 ^e	<1	<1
Congenital abnormalities	20,000–30,000	10	10–100
Other disorders of complex etiology ^f			
Heart disease	600,000	NE	NE
Cancer	300,000	NE	NE
Selected others	300,000	NE	NE

Note: VSI, very slow increase; VLI, very little increase; NE, not estimated.

^aRisks pertain to average population exposure of 10 mSv (1 rem) per generation to a population with the spontaneous genetic burden of humans and a doubling dose for chronic exposure of 1 Sv (100 rem).

^b"Clinically severe" assumes that survival and reproduction are reduced by 20–80% relative to normal.

^c"Clinically mild" assumes that survival and reproduction are reduced by 1–20% relative to normal.

^dObtained by subtracting an estimated 2,500 clinically severe dominant traits from an estimated total incidence of dominant traits of 10,000.

^eEstimated frequency from UNSCEAR Reports (1982 and 1986).

^fMost frequent result of chromosomal nondisjunction among liveborn children. Estimated frequency from UNSCEAR (1982 and 1986).

^gLifetime prevalence estimates may vary according to diagnostic criteria and other factors. The values given for heart disease and cancer are round-numbered approximations for all varieties of the diseases. With regard to heart disease, no implication is made that any form of heart disease is caused by radiation among exposed individuals. The effect, if any, results from mutations that may be induced by radiation and expressed in later generations, which contribute, along with other genes, to the genetic component of susceptibility. This is analogous to environmental risk factors that contribute to the environmental component of susceptibility. The magnitude of the genetic component in susceptibility to heart disease and other disorders with complex etiologies is unknown.

Source: Adapted from the 1990 National Academy of Sciences/National Research Council Committee on the Biological Effects of Ionizing Radiation. *The Health Effects of Exposure to Low Levels of Ionizing Radiation (BEIR V)*. Washington, DC: NAS/NRC, 1990.

births in the succeeding generation, as a result of increases in the number of autosomal dominant and (to a much lesser degree) sex-linked dominant mutations (Table 25-10). If a population is continually exposed to an increased radiation dose of 10 mGy (1 rad) for each generation, an equilibrium will be established between the induction of new genetic disorders and the loss of existing ones. At equilibrium, an additional 100 genetic disorders would be expected in the population, with the majority contributed by clinically mild autosomal dominant mutations. Chromosomal damage and recessive mutations are thought to make only minor contributions to the equilibrium rate. Fortunately, serious chromosomal damage is either lethal to the cell or is selected out, and recessive mutations require both homologous alleles for expression. Table 25-10 also lists other disorders with complex etiologies, such as cancer and heart disease. Although some cancers have genetic origins, the ability of radiation to influence the inheritance of cancer susceptibility has not been demonstrated.

Through natural selection, the gene pool has the capacity to absorb large amounts of radiation damage without significantly affecting the population. Variations in exposure to natural background do not contribute significantly to a population's genetic risk. Even a dose of 100 mGy (10 rad) would produce only about 200 additional genetic disorders per 1 million live births in the first generation (0.2%/Gy or 0.002%/rad), compared with the normal incidence of approximately 1 in 20 or 5% (some estimates are 1 in 10). Therefore, the 100-mGy (10-rad) dose would cause an increase in the spontaneous rate of genetic disorders of less than 0.4%.

Typical diagnostic and occupational radiation exposures, although increasing the dose to the gonads of those exposed, would not be expected to result in any significant genetic risk to their progeny. Therefore, although delaying conception after therapeutic doses of radiation to reduce the probability of transmission of genetic damage to offspring is effective, it is not common practice in diagnostic imaging.

25.7 RADIATION EFFECTS IN UTERO

Developing organisms are highly dynamic systems that are characterized by rapid cell proliferation, migration, and differentiation. Thus, the developing embryo is extremely sensitive to ionizing radiation, as would be expected based on Bergonié and Tribondeau's laws of radiosensitivity. The response after exposure to ionizing radiation depends on a number of factors including (a) total dose, (b) dose rate, (c) radiation quality, and (d) the stage of development at the time of exposure. Together, these factors determine the type and extent of the damage that would be of concern after an exposure, among which are prenatal or neonatal death, congenital abnormalities, growth impairment, reduced intelligence, genetic aberrations, and an increase in risk of cancer.

Radiation Effects and Gestation

The gestational period can be divided into three stages: a relatively short *preimplantation* stage, followed by an extended period of *major organogenesis*, and finally the *fetal growth* stage, during which differentiation is complete and growth mainly occurs. Each of these stages is characterized by different responses to radiation expo-

sure, owing principally to the relative radiosensitivities of the tissues at the time of exposure.

Preimplantation

The preimplantation stage begins with the union of the sperm and egg and continues through day 9 in humans, when the zygote becomes embedded in the uterine wall. During this period, the two pronuclei fuse, cleave, and form the morula and blastula.

The conceptus is extremely sensitive during the preimplantation stage and radiation damage can result in prenatal death. During this period the incidence of congenital abnormalities is low, although not completely absent. Embryos exhibit the so-called *all-or-nothing response*, in which, if prenatal death does not occur, the damaged cells are repaired or replaced to the extent that there are unlikely to be visible signs of abnormalities even though radiation may have killed several cells.

Several factors, including repair capability, lack of cellular differentiation, and the relatively hypoxic state of the embryo, are thought to contribute to its resistance to radiation-induced abnormalities. During the first few divisions, the cells are undifferentiated and lack predetermination for a particular organ system. If radiation exposure were to kill some cells at this stage, the remaining cells could continue the embryonic development without gross malformations because they are indeterminate. However, chromosomal damage at this point may be passed on and expressed at some later time. When cells are no longer indeterminate, loss of even a few cells may lead to anomalies, growth retardation, or prenatal death. The most sensitive times of exposure in humans are at 12 hours after conception, when the two pronuclei fuse to the one-cell stage, and again at 30 and 60 hours when the first two divisions occur.

Chromosomal aberrations from radiation exposure at the one-cell stage could result in loss of a chromosome in subsequent cell divisions that would then be uniform throughout the embryo. Most chromosomal loss at this early stage is lethal. Loss of a sex chromosome in female embryos may produce Turner's syndrome.

The woman may not know she is pregnant during the preimplantation period, the time at which the conceptus is at greatest risk of lethal effects. Animal experiments have demonstrated an increase in the spontaneous abortion rate after doses as low as 50 to 100 mGy (5 to 10 rad) delivered during the preimplantation period. After implantation, doses in excess of 250 mGy (25 rad) are required to induce prenatal death. The spontaneous abortion rate has been reported to be between 30% and 50%.

Organogenesis

Embryonic malformations occur more frequently during the period of major organogenesis (2nd to 8th week after conception). The initial differentiation of cells to form certain organ systems typically occurs on a specific gestational day. For example, neuroblasts (stem cells of the CNS) appear on the 18th gestational day, the forebrain and eyes begin to form on day 20, and primitive germ cells are evident on day 21. Each organ system is not at equal risk during the entire period of major organogenesis. In general, the greatest probability of a malformation in

a specific organ system (the so-called *critical period*) exists when the radiation exposure is received during the period of peak differentiation of that system. This may not always be the case, however, because damage can occur to adjacent tissue, which has a negative effect on a developing organ system. Some anomalies may have more than one critical period. For example, cataract formation has been shown to have three critical periods in mice. Figure 25-19 shows the critical periods in human fetal development for various radiation-induced birth defects.

The only organ system (in humans or laboratory rodents) that has shown an association between malformations and low-LET radiation doses less than 250 mGy (25 rad) is the CNS. Embryos exposed early in organogenesis exhibit the greatest intrauterine growth retardation, presumably because of cell depletion. *In utero* exposure to doses greater than 100 mGy (10 rad) of mixed neutron and gamma radiation from the Hiroshima atomic bomb resulted in a significant increase in the incidence of microcephaly. Low doses of x-rays have also been shown to produce growth retardation and behavioral defects.

In general, radiation-induced teratogenic effects are less common in humans than in animals. This is primarily because in humans a smaller fraction of the gestational period is taken up by organogenesis (about $1/15$ of the total, compared with

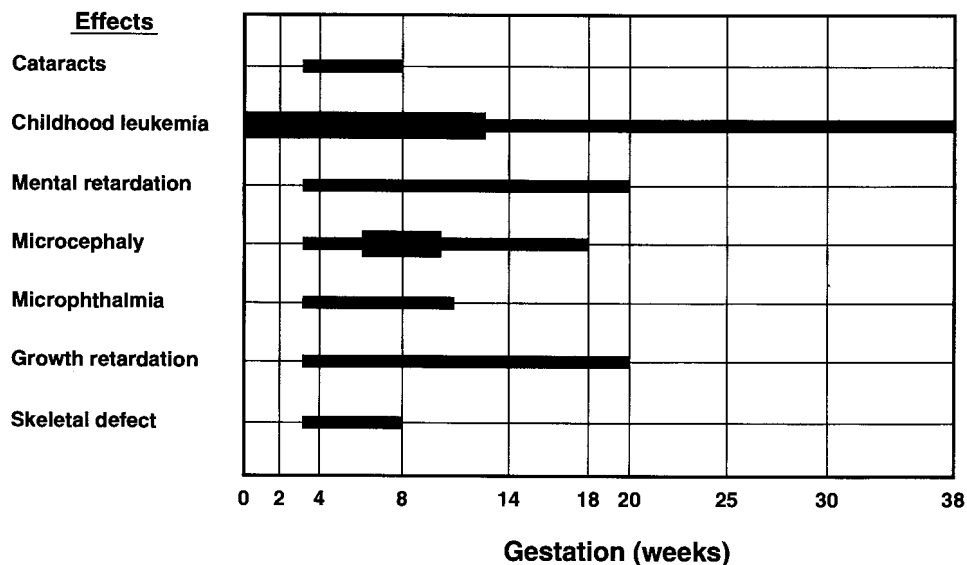


FIGURE 25-19. Critical periods for radiation-induced birth defects in humans. Data were obtained from children who were exposed *in utero* as a result of medical radiation treatments of their mothers. Fetal doses were in excess of 2.5 Gy (250 rad). Most children had more than one anomaly, and mental retardation was usually associated with microcephaly when exposure occurred during this period of gestation. No significant increase in any of the effects listed here have been shown to occur in humans before the 7th week of gestation. The radiation-induced risks reported before the 7th week of gestation are from animal data. The broader lines indicate increased probability of effect. (Data from Dekaban AS. Abnormalities in children exposed to x-irradiation during various stages of gestation: tentative time table of radiation injury to human fetus. *J Nucl Med* 1968;9:471-477.)

$\frac{1}{3}$ for mice). Nevertheless, development of the CNS in humans takes place over a much longer gestational interval than in experimental animals, and therefore the CNS is more likely to be a target for radiation induced-damage. An important distinguishing characteristic of teratogenic effects in humans is the concomitant effects on the CNS and/or fetal growth. All cases of human *in utero* irradiation that have resulted in gross malformations have been accompanied by CNS abnormalities or growth retardation or both.

The response of each organ to the induction of radiation-induced malformations is unique. Such factors as gestational age; radiation quantity, quality, and dose rate; oxygen tension; the cell type undergoing differentiation and its relationship to surrounding tissues; and other factors influence the outcome.

Fetal Growth Stage

The fetal growth stage in humans begins after the end of major organogenesis (day 45) and continues until term. During this period the incidence of radiation-induced prenatal death and congenital anomalies is, for the most part, negligible. Anomalies of the nervous system and sense organs are the primary radiation-induced abnormalities observed during this period, which coincides with their relative growth and development. Much of the damage induced at the fetal growth stage may not be manifested until later in life as behavioral alterations or reduced intelligence (e.g., IQ).

Human Irradiation In Utero

Two groups that have been studied for the effects of *in utero* irradiation are the children of the atomic-bomb survivors and children whose mothers received medical irradiation (diagnostic and/or therapeutic) during pregnancy. The predominant effects that were observed included microcephaly and mental and growth retardation. Eye, genital, and skeletal abnormalities occurred less frequently. Excluding the mentally retarded, individuals exposed *in utero* at Hiroshima and Nagasaki between the 8th to 25th week after conception period demonstrated poorer IQ scores and school performance than did unexposed children. No such effect was seen in those exposed before the 8th week or after the 25th week. The decrease in IQ was dose dependent, with an apparent threshold below 250 mGy (25 rad). The greatest sensitivity for radiation-induced mental retardation is seen between the 8th and 15th weeks, during which the risk is approximately 1/250 or 0.40% per Gy (100 rad) to the fetus.

Microcephaly was observed in children exposed *in utero* at the time of the atomic bomb detonation. For those exposed before the 18th gestational week, the incidence of microcephaly was proportional to dose, up to 1.5 Gy (150 rad), above which the decreased incidence was presumably due to the increase in fetal mortality. The incidence of microcephaly (normally ~3%) was increased when exposures occurred during the first or second trimester. No excess was found with exposure in the third trimester, regardless of dose. The incidence of microcephaly in the fetal dose range of 100 to 490 mGy (10 to 49 rad) was approximately 19% and 6% for the first and second trimester, respectively. At fetal doses greater than 1 Gy (100 rad), the incidence of microcephaly rose to approximately 83% and 42%, for exposures in the first and second trimester, respectively.

Similar effects have been reported after *in utero* exposure during medical irradiation of the mother. Twenty of 28 children irradiated *in utero* as a result of pelvic radium or x-ray therapy were mentally retarded, among which 16 were also microcephalic. Other deformities, including abnormal appendages, hydrocephaly, spina bifida, blindness, cataracts, and microphthalmia have been observed in humans exposed *in utero*.

Although each exposure needs to be evaluated individually, the prevailing scientific opinion is that there are thresholds for most congenital abnormalities. Doses lower than 100 mGy (10 rad) are generally thought to carry negligible risk compared with the spontaneous incidence of congenital anomalies (4% to 6%). Therapeutic abortion below this threshold would not usually be justified. At one time, radiation was widely used to induce therapeutic abortion in cases where surgery was deemed inadvisable. The standard treatment was 3.5 to 5 Gy (350 to 500 rad) given over 2 days, which typically resulted in fetal death within 1 month. In most instances, fetal irradiation from diagnostic procedures rarely exceeds 50 mGy (5 rad), which has not been shown to place the fetus at any significant increased risk for congenital malformation or growth retardation.

Cancer Incidence after *In Utero* Irradiation

A correlation between childhood leukemia and solid tumors and *in utero* diagnostic radiation was reported by Stewart in 1956 in a retrospective study of childhood cancer in Great Britain (often referred to as the “Oxford Survey of Childhood Cancers”). This observation was confirmed by similar studies reporting a relative risk of approximately 1.4 for childhood cancer; however, the effect was not observed among survivors of the atomic bomb who were irradiated *in utero*. Absence of a positive finding in the atomic-bomb survivors could have been predicted on a statistical basis, owing to the relatively small sample size. These positive studies have been criticized based on a variety of factors, including the influence of the pre-existing medical condition for which the examination was required and the lack of good estimates of fetal doses. More recently however, Doll and Wakeford reviewed the scientific evidence and concluded that irradiation of the fetus *in utero* (particularly in the last trimester) does increase the risk of childhood cancer, that the increase in risk is produced by doses on the order of 10 mGy (1 rad), and that the excess risk is approximately 6% per Gy.

By way of comparison, the natural total risk of mortality from malignancy through age 10 years is approximately 1/1,200. If a chest x-ray delivered 0.6 μ Gy (60 μ rad) to the fetus, the probability of development of a fatal cancer during childhood from the exposure would be less than 1 in 27 million! As indicated in Chapter 23, regulatory agencies limit occupational radiation exposure to the fetus to no more than 5 mSv (500 mrem) during the gestational period, provided that the dose rate does not substantially exceed 0.5 mSv (50 mrem) in any one month. If the maximally allowed dose were received, the probability of developing a fatal cancer during childhood from the exposure would be approximately 1 in 3,300. Table 25-11 presents a comparison of the risks of radiation exposure with other risks encountered during pregnancy. It is clear from these data that, although radiation exposure should be minimized, its risk at levels associated with occupational and diagnostic exposures is nominal when compared with other potential hazards.

TABLE 25-11. EFFECT OF RISK FACTORS ON PREGNANCY-OUTCOME

Effect	Number occurring from natural causes	Risk factor	Excess occurrences from risk factors
Radiation Risks			
Childhood Cancer			
Cancer death in children	1.4/1,000	Radiation dose of 10 mSv (1 rem) received before birth	0.6/1,000
Abnormalities		Radiation dose of 10 mSv (1 rem) received during specific periods after conception:	
Small head size	40/1,000	4–7 wk	5/1,000
Small head size	40/1,000	8–11 wk	9/1,000
Mental retardation	4/1,000	8–15 wk	4/1,000
Nonradiation Risks			
Occupation			
Stillbirth or spontaneous abortion	200/1,000	Work in high-risk occupations	90/1,000
Alcohol Consumption			
Fetal alcohol syndrome	1–2/1,000 ^a	2–4 drinks/day	100/1,000
Fetal alcohol syndrome	1–2/1,000 ^a	More than 4 drinks/day	200/1,000
Fetal alcohol syndrome	1–2/1,000 ^a	Chronic alcoholic (>10 drinks/day)	350/1,000
Perinatal infant death	23/1,000	Chronic alcoholic (>10 drinks/day)	170/1,000
Smoking			
Perinatal infant death	23/1,000	<1 pack/day	5/1,000
Perinatal infant death	23/1,000	≥1 pack/day	10/1,000

^aThere is a naturally occurring syndrome that has the same symptoms as a full-blown fetal alcohol syndrome that occurs in children born to mothers who have not consumed alcohol.

Source: U.S. Nuclear Regulatory Commission. *Instruction concerning prenatal radiation exposure*. Regulatory Guide 8.13, Rev. 2.

Risks from *In Utero* Exposure to Radionuclides

Radiopharmaceuticals may be administered to pregnant women if the diagnostic information to be obtained outweighs the risks associated with the radiation dose. The principal risk in this regard is the dose to the fetus. Radiopharmaceuticals can be divided into two broad categories: those that cross the placenta and those that remain on the maternal side of the circulation. Radiopharmaceuticals that do not cross the placenta irradiate the fetus by the emission of penetrating radiation (mainly gamma rays and x-rays). Early in the pregnancy, when the embryo is small and the radiopharmaceutical distribution is fairly uniform, the dose to the fetus can be approximated by the dose to the ovaries. Estimates of early-pregnancy fetal doses from commonly used radiopharmaceuticals are listed in Appendix D-4. Radiopharmaceuticals that cross the placenta may be distributed in the body of the fetus or be concentrated locally if the fetal target organ is mature enough. A classic (and important) example of such a radiopharmaceutical is radioiodine.

Radioiodine rapidly crosses the placenta, and the fetal thyroid begins to concentrate radioiodine after the 13th week of gestation. Between 14 and 22 weeks of gesta-

TABLE 25-12. ABSORBED DOSE mGy/MBq (rad/mCi) IN FETAL THYROID FROM RADIONUCLIDES GIVEN ORALLY TO MOTHER

Fetal Age (mo)	Iodine-123	Iodine-125	Iodine-131	Technetium-99m
3	2.7 (10)	290 (1,073)	230 (851)	0.032 (0.12)
4	2.6 (10)	240 (888)	260 (962)	
5	6.4 (24)	280 (1,036)	580 (2,146)	
6	6.4 (24)	210 (777)	550 (2,035)	0.15 (0.54)
7	4.1 (15)	160 (592)	390 (1,443)	
8	4.0 (15)	150 (555)	350 (1,295)	
9	2.9 (11)	120 (444)	270 (1,000)	0.38 (1.40)

Source: Watson EE. Radiation Absorbed Dose to the Human Fetal Thyroid. In: Watson EE, Schlaske-Stelson A, eds. *5th International Radiopharmaceutical Dosimetry Symposium*. Oak Ridge, TN. May 7–10, 1991. Oak Ridge, TN: Oak Ridge Associated Universities, 1992.

tion, the percentage uptake of radioiodine by the fetal thyroid exceeds that of the adult, ranging from 55% to 75%. The biologic effect on the fetal thyroid is dose dependent and can result in hypothyroidism or ablation if the activity administered to the mother is high enough. For I-131 estimates of dose to the fetal thyroid range from 230 to a maximum of 580 mGy/MBq (851 to 2,146 rad/mCi) for gestational ages between 3 and 5 months. Total body fetal dose ranges from 0.072 mGy/MBq (0.266 rad/mCi) early in pregnancy to a maximum of 0.27 mGy/MBq (1 rad/mCi) near term. Table 25-12 presents the fetal thyroid dose from various radioiodines and Tc-99m.

Summary

Radiation exposure of the embryo at the preimplantation stage usually leads to an all-or-nothing phenomenon (i.e., either fetal death and resorption or normal fetal risk). During the period of organogenesis, the risk of fetal death decreases substantially, whereas the risk of congenital malformation coincides with the peak developmental periods of various organ systems. Exposures in excess of 1 Gy (100 rad) are associated with a high incidence of CNS abnormalities. During the fetal growth stage *in utero*, exposure poses little risk of congenital malformations; however, growth retardation, abnormalities of the nervous system, and the risk of childhood cancer can be significant. Growth retardation after *in utero* exposure as low as 100 mGy (10 rad) has been demonstrated. Fetal doses from most diagnostic x-ray and nuclear medicine examinations are lower than 50 mGy (5 rad) and have not been demonstrated to produce any significant impact on fetal growth and development. A number of epidemiologic studies have evaluated the ability of *in utero* radiation exposure to increase the risk of childhood neoplasms. Although these studies show conflicting results and remain controversial, a conservative estimate of the excess risk of childhood cancer from *in utero* irradiation is approximately 6% per Gy (0.06% per rad). Figure 25-20 illustrates the relative incidence of radiation-induced health effects at various stages in fetal development.

The risk of *in utero* radiation exposure can be placed into perspective by considering the probability of birthing a healthy child after a given dose to the conceptus (Table 25-13). Clearly, even a conceptus dose of 10 mSv (1 rem) does not significantly affect the risks associated with pregnancy. Fetal doses from diagnostic

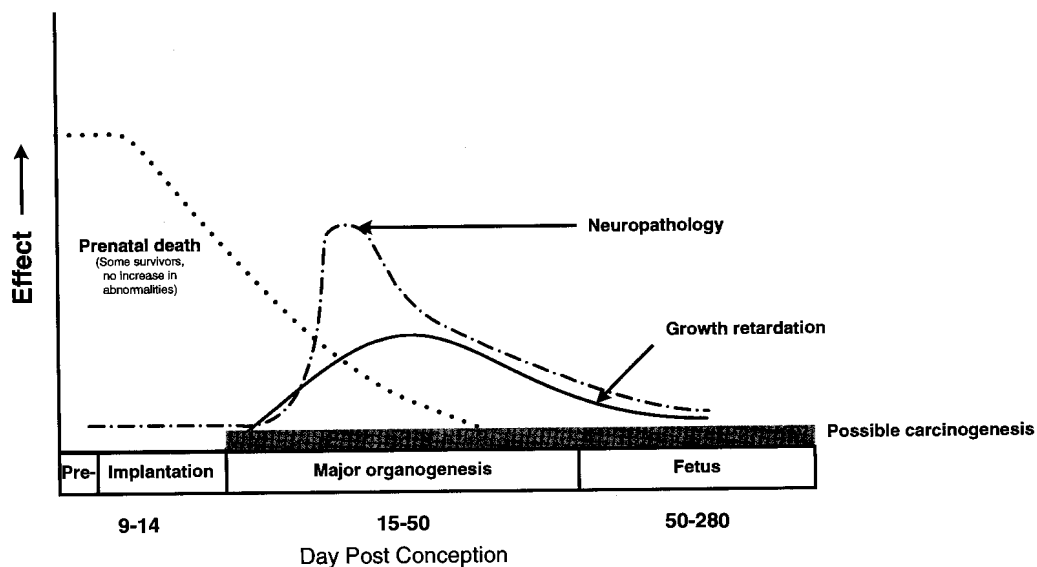


FIGURE 25-20. Relative incidence of various adverse health effects associated with radiation exposure *in utero* at various stages of gestation. (Adapted from Mettler FA, Upton AC. *Medical effects of ionizing radiation*, 2nd ed. Philadelphia: WB Saunders, Co., 1995.

examinations rarely justify therapeutic abortion; however, each case should be evaluated individually and the risks should be explained to the patient. Every effort should be made to reduce radiation exposure to the patient and especially to the fetus. However, considering the relatively small risk associated with diagnostic examinations, postponing clinically indicated examinations or scheduling examinations around the patient's menstrual cycle to avoid irradiating a potential conceptus are unwarranted measures. Nevertheless, every fertile female patient should be asked whether she might be pregnant before diagnostic examinations or therapeutic procedures using ionizing radiation are performed. If the patient is pregnant and alternative diagnostic or therapeutic procedures are inappropriate, the risks and benefits of the procedure should be discussed with the patient.

TABLE 25-13. PROBABILITY OF BIRTHING HEALTHY CHILDREN

Dose ^a to Conceptus (mSv [mrem])	Child with No Malformation (Percentage)	Child Will Not Develop Cancer (Percentage)	Child Will Not Develop Cancer or Have a Malformation (Percentage)
0 (0)	96	99.93	95.93
0.5 (50)	95.999	99.927	95.928
1.0 (100)	95.998	99.921	95.922
2.5 (250)	95.995	99.908	95.91
5.0 (500)	95.99	99.89	95.88
10.00 (1,000)	95.98	99.84	95.83

^aRefers to absorbed dose above natural background. This table assumes conservative risk estimates, and it is possible that there is no added risk.

Source: From Wagner LK, Hayman LA. Pregnancy in women radiologists. *Radiology* 1982;145:559-562.

SUGGESTED READING

- Doll R, Wakeford F. Risk of childhood cancer from fetal irradiation. *Br J Radiol* 1997;70: 130–139.
- Hall EJ. *Radiobiology for the radiologist*, 5th ed. Philadelphia: Lippincott Williams & Wilkins, 2000.
- Mettler FA, Upton AC. *Medical effects of ionizing radiation*, 2nd ed. Philadelphia: WB Saunders, 1995.
- National Research Council, Committee on the Biological Effects of Ionizing Radiations. *Health effects of exposure to low levels of ionizing radiation (BEIR V)*. Washington, DC: National Academy Press, 1990.
- Wagner LK, et al. *Radiation bioeffects and management test and syllabus*. American College of Radiology Professional Self-Evaluation and Continuing Education Program, Number 32. Reston, VA: American College of Radiology, 1991.
- Wagner LK, Lester RG, Saldana LR. *Exposure of the pregnant patient to diagnostic radiations: a guide to medical management*, 2nd ed. Madison, WI: Medical Physics Publishing, 1997.

SECTION
V

APPENDICES

Appendix A

FUNDAMENTAL PRINCIPLES OF PHYSICS

A.1 PHYSICS LAWS, QUANTITIES, AND UNITS

Laws of Physics

Physics is the study of the physical environment around us—from the smallest quarks to the galactic dimensions of black holes and quasars. Much of what is known about physics can be summarized in a set of laws that describe physical reality. These laws are based on reproducible results from physical observations that are consistent with theoretical predictions. A well-conceived law of physics is applicable in a wide range of circumstances. A physical law that states that the force exerted by gravity on an individual on the surface of the Earth is 680 newtons is a poor example, because it is a description of a single situation only. Newton's Law of Gravity, however, describes the gravitational forces between any two bodies at any distance from each other and is an example of a well-conceived, generally usable law.

Vector versus Scalar Quantities

For some quantities, such as force, velocity, acceleration, and momentum, direction is important, in addition to the magnitude. These quantities are called *vector* quantities. Quantities that do not incorporate direction, such as mass, time, energy, electrical charge, and temperature, are called *scalar* quantities.

A vector quantity is represented graphically by an arrow whose length is proportional to the magnitude of the vector. A vector quantity is represented by bold-face type, as in the equation $\mathbf{F} = m\mathbf{a}$, where $\mathbf{F}$ and $\mathbf{a}$ are vector quantities and m is a scalar quantity.

In many equations of physics, such as $\mathbf{F} = m\mathbf{a}$, a vector is multiplied by a scalar. In vector-scalar multiplication, the magnitude of the resultant vector is the product of the magnitude of the original vector and the scalar; however, the direction of the vector is not changed by the multiplication. If the scalar has units, the multiplication may also change the units of the vector. Force, for example, has different units than acceleration. Vector-scalar multiplication is shown in Fig. A-1.

Two or more vectors may be added. This addition is performed graphically by placing the tail of one vector against the head of the other, as shown in Fig. A-2. The resultant vector is that reaching from the tail of the first to the head of the second. The order in which vectors are added does not affect the result. A special case occurs when one vector has the same magnitude but opposite direction of the other; in this case, the two vectors cancel, resulting in no vector.

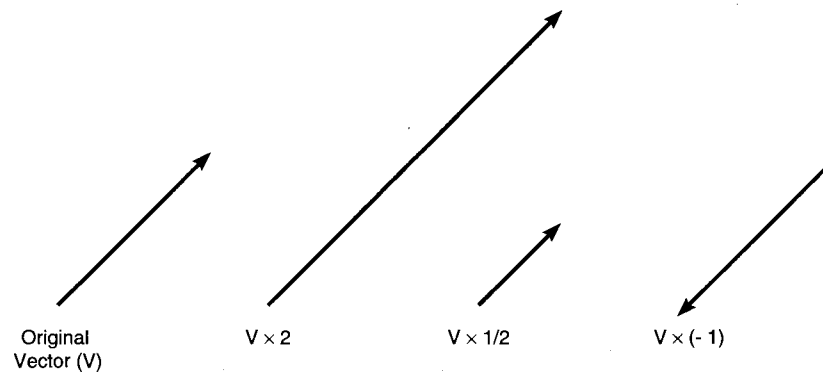


FIGURE A-1. Vector-scalar multiplication.

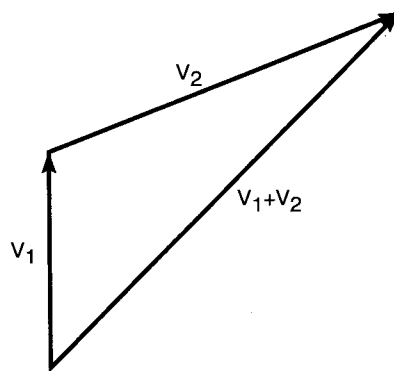


FIGURE A-2. Vector addition.

International System of Units

As science has developed, many disciplines have devised their own specialized units for measurements. An attempt has been made to establish a single set of units to be used across all disciplines of science. This set of units, called the *Système International* (SI), is gradually replacing the traditional units. The SI establishes a set of seven base units—the kilogram, meter, second, ampere, kelvin, candela, and mole—whose magnitudes are carefully described by standards laboratories. Two supplemental units, the radian and steradian, are used to describe angles in two and three dimensions, respectively. All other units, called *derived units*, are defined in terms of these fundamental units. For representing quantities much greater than or much smaller than an SI unit, the SI unit may be modified by a prefix. Commonly used prefixes are shown in Table A-1.

TABLE A-1. COMMONLY USED PREFIXES FOR UNITS

Prefix	Meaning	Prefix	Meaning
centi (c)	10^{-2}	kilo (k)	10^{+3}
milli (m)	10^{-3}	mega (M)	10^{+6}
micro (μ)	10^{-6}	giga (G)	10^{+9}
nano (n)	10^{-9}	tera (T)	10^{+12}
pico (p)	10^{-12}		

A.2 CLASSICAL PHYSICS

Mass, Length, and Time

Mass is a measure of the resistance a body has to acceleration. It should not be confused with weight, which is the gravitational force exerted on a mass. The SI unit for mass is the kilogram (kg). The SI unit for length is the meter (m) and for time is the second (s).

Velocity and Acceleration

Velocity, a vector quantity, is the rate of change of position with respect to time:

$$v = \Delta x / \Delta t \quad [A-1]$$

The magnitude of the velocity, called the *speed*, is a scalar quantity. The SI unit for velocity and speed is the meter per second (m/s). *Acceleration*, also a vector quantity, is defined as the rate of change of velocity with respect to time:

$$a = \Delta v / \Delta t \quad [A-2]$$

The SI unit for acceleration is the meter per second per second (m/s²).

Forces and Newton's Laws

Force, a vector quantity, is a push or pull. The physicist and mathematician Isaac Newton proposed three laws regarding force and velocity:

1. Unless acted upon by external force, an object at rest remains at rest and an object in motion remains in uniform motion. That is, an object's velocity remains unchanged unless an external force acts upon the object.
2. For every action, there is an equal and opposite reaction. That is, if one object exerts a force on a second object, the second object also exerts a force on the first that is equal in magnitude but opposite in direction.
3. A force acting upon an object produces an acceleration in the direction of the applied force:

$$F = ma \quad [A-3]$$

The SI unit for force is the newton (N): 1 N is defined as 1 kg-m/s². There are four types of forces: *gravitational*, *electrical*, *magnetic*, and *nuclear*. These forces are discussed later.

Energy

Kinetic Energy

Kinetic energy, a scalar quantity, is a property of moving matter that is defined by the following equation:

$$E_k = \frac{1}{2}mv^2 \quad [A-4]$$

where E_k is the kinetic energy, m is the mass of the object, and v is the speed of the object. (Einstein discovered a more complicated expression that must be used for objects with speeds approaching the speed of light.) The SI unit for all forms of energy is the joule (J): 1 J is defined as 1 kg-m²/s².

Potential Energy

Potential energy is a property of an object in a force field. The force field may be a gravitational force field. If an object is electrically charged, it may be an electrical force field caused by nearby charged objects. If an object is very close to an atomic nucleus, it may be influenced by nuclear forces. The potential energy (E_p) is a function of the position of the object; if the object changes position with respect to the force field, its potential energy changes.

Conservation of Energy

The total energy of an object is the sum of its kinetic and potential energies:

$$E_{\text{total}} = E_k + E_p \quad [\text{A-5}]$$

Aside from friction, the total energy of the object does not change. For example, consider a brick held several feet above the ground. It has a certain amount of gravitational potential energy by virtue of its height, but it has no kinetic energy because it is at rest. When released, it falls downward with a continuous increase in kinetic energy. As the brick falls, its position changes and its potential energy decreases. The sum of its kinetic and potential energies does not change during the fall. One may think of the situation as one in which potential energy is converted into kinetic energy.

Momentum

Momentum, like kinetic energy, is a property of moving objects. However, unlike kinetic energy, momentum is a vector quantity. The momentum of an object is defined as follows:

$$\mathbf{p} = m\mathbf{v} \quad [\text{A-6}]$$

where $\mathbf{p}$ is the momentum of an object, m is its mass, and $\mathbf{v}$ is its velocity. The momentum of an object has the same direction as its velocity. The SI unit of momentum is the kilogram-meter per second (kg-m/s). It can be shown from Newton's Laws that, if no external forces act on a collection of objects, the total momentum of the set of objects does not change. This principle is called the Law of Conservation of Momentum.

A.3 ELECTRICITY AND MAGNETISM

Electricity

Electrical Charge

Electrical charge is a property of matter. Matter may have a positive charge, a negative charge, or no charge. The charge on an object may be determined by observing how it behaves in relation to other charged objects. The SI unit of charge is the *coulomb* (C). The smallest magnitude of charge is that of the electron. There are approximately 10^{19} electron charges per coulomb.

A fundamental law of physics states that electrical charge is conserved. This means that, if charge is neither added to nor removed from a system, the total

amount of charge in the system does not change. The signs of the charges must be considered in calculating the total amount of charge in a system. If a system contains two charges of the same magnitude but of opposite sign—for example, in a hydrogen atom, in which a single electron (negative charge) orbits a single proton (positive charge)—the total charge of the system is zero.

Electrical Forces and Fields

The force (F) exerted on a charged particle by a second charged particle is described by the equation

$$F = kq_1q_2/r^2 \quad [A-7]$$

where q_1 and q_2 are the charges on the two particles, r is the distance between the two particles, and k is a constant. The direction of the force on each particle is either toward or away from the other particle; the force is attractive if one charge is negative and the other positive, and it is repulsive if both charges are of the same sign.

An electrical field exists in the vicinity of electrical charges. To measure the electrical field strength at a point, a small test charge is placed at that point and the force on the test charge is measured. The electrical field strength (E) is then calculated as follows:

$$E = F/q \quad [A-8]$$

where F is the force exerted on the test charge and q is the magnitude of the test charge. Electrical field strength is a vector quantity. If the test charge q is positive, E and F have the same direction; if the test charge is negative, E and F have opposite directions.

An electrical field may be visualized as lines of electrical field strength existing between electrically charged entities. The strength of the electrical field is depicted as the density of these lines. Figure A-3 shows the electrical fields surrounding a point charge and between two parallel plate electrodes.

An *electrical dipole* has no net charge but does possess regional distribution of positive and negative charge. When placed in an electrical field, the dipole tends to align with the field because of the torque exerted on it by that field (attraction by unlike charges, repulsion by like charges). The electrical dipole structure of certain natural and man-made crystals is used in ultrasound imaging devices to produce and detect sound waves, are described in Chapter 16.

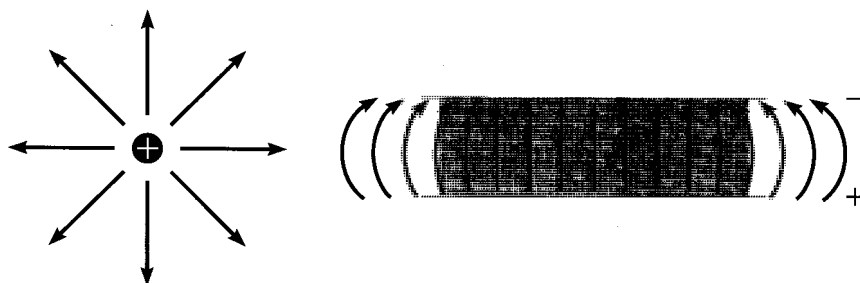


FIGURE A-3. Electrical fields around a point charge (**left**) and between two parallel charged electrodes (**right**).

Electrical Forces on Charged Particles

When a charged particle is placed in an electrical field, it experiences a force that is equal to the product of its charge q and the electrical field strength E at that location:

$$F = qE$$

If the charge on the particle is positive, the force is in the direction of the electrical field at that position; if the charge is negative, the force is opposite to the electrical field. If the particle is not restrained, it experiences an acceleration:

$$a = qE/m$$

where m is the mass of the particle.

Electrical Current

Electrical current is the flow of electrical charge. Charge may flow through a solid, such as a wire; through a liquid, such as the acid in an automobile battery; through a gas, as in a fluorescent light; or through a vacuum, as in an x-ray tube. Current may be the flow of positive charges, such as positive ions or protons in a cyclotron, or it may be the flow of negative charges, such as electrons or negative ions. In some cases, such as an ionization chamber radiation detector, there is a flow of positive charges in one direction and a flow of negative charges in the opposite direction. Although current is often the flow of electrons, the direction of the current is defined as the flow of positive charge (Fig. A-4). The SI unit of current is the *ampere* (A): 1 A is 1 C of charge crossing a defined boundary per second.

Electrical Potential Difference

Electrical potential, more commonly called *voltage*, is a scalar quantity. Voltage is the difference in the electrical potential energy of an electrical charge at two positions, divided by the charge:

$$V_{ab} = (E_{pa} - E_{pb})/q \quad [A-9]$$

where E_{pa} is the electrical potential energy of the charge at location a , E_{pb} is the electrical potential energy of the same charge at location b , and q is the amount of charge. It is meaningless to specify the voltage at a single point in an electrical circuit; it must be specified in relation to another point in the circuit. The SI unit of electrical potential is the volt (V): 1 V is defined as 1 joule per coulomb, or $1 \text{ kg} \cdot \text{m}^2/\text{s}^2 \cdot \text{C}$.

Potential is especially useful in determining the final kinetic energy of a charged particle moving between two electrodes through a vacuum, as in an x-ray tube. According to the principle of conservation of energy (Equation A-5), the gain in kinetic energy of the charged particle is equal to its loss of potential energy:

$$E_{k\text{-final}} - E_{k\text{-initial}} = E_{p\text{-initial}} - E_{p\text{-final}} \quad [A-10]$$

Assuming that the charged particle starts with no kinetic energy ($E_{k\text{-initial}} = 0$) and using Equation A-9, Equation A-10 becomes:

$$E_{k\text{-final}} = qV \quad [A-11]$$

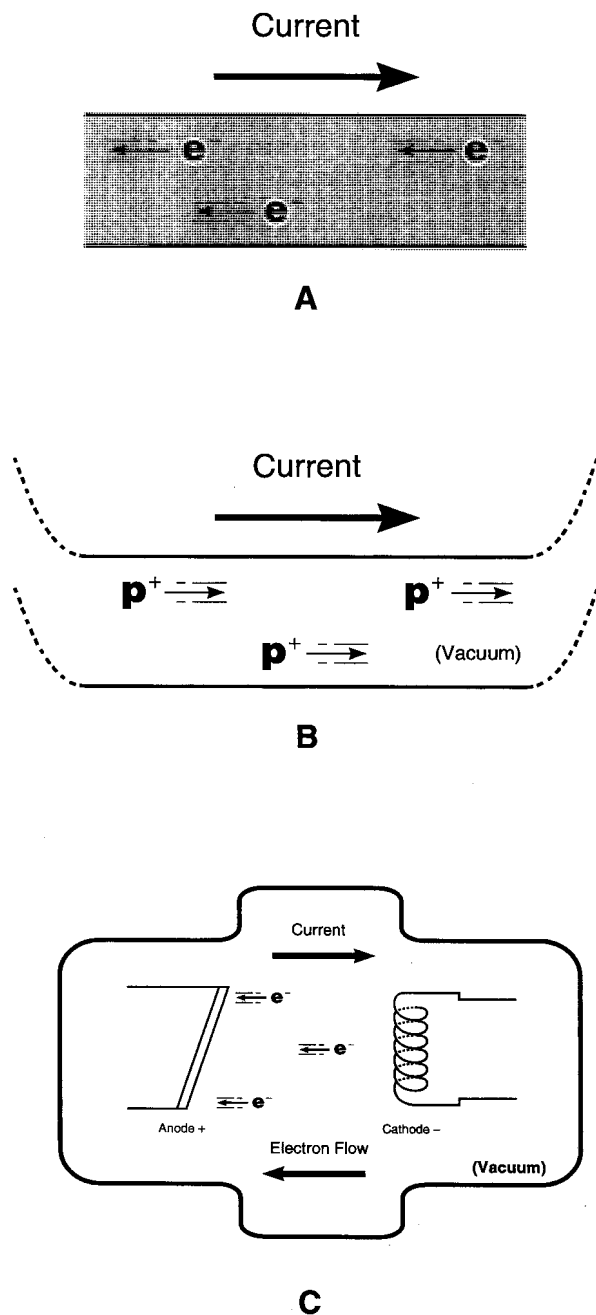


FIGURE A-4. Direction of electrical current. **A:** A situation is shown in which electrons are flowing to the left. **B:** Positive charges are shown moving to the right. In both cases, the direction of current is to the right. **C:** Although the electrons in an x-ray tube pass from cathode to anode, current technically flows in the opposite direction.

where q is the charge of the particle and V is the potential difference between the two electrodes.

For example, the final kinetic energy of an electron (charge = 1.602×10^{-19} C) accelerated through an electrical potential difference of 100 kilovolts is

$$E_{k\text{-final}} = qV = (1.602 \times 10^{-19} \text{ C}) (100 \text{ kV}) = 1.602 \times 10^{-14} \text{ J}$$

The joule is a rather large unit of energy for subatomic particles. In atomic and nuclear physics, energies are often expressed in terms of the *electron volt* (eV). One electron volt is the kinetic energy developed by an electron accelerated across a potential difference of 1 V. One electron volt is equal to 1.602×10^{-19} J.

For example, the kinetic energy of an electron, initially at rest, which is accelerated through a potential difference of 100 kilovolts is

$$E_k = qV = (1 \text{ electron charge})(100 \text{ kV}) = 100 \text{ keV}$$

One must be careful not to confuse the units of potential difference (V, kV, and MV) with units of energy (eV, keV, and MeV).

Electrical Power

Power is defined as the rate of the conversion of energy from one form to another with respect to time:

$$P = \Delta E / \Delta t \quad [\text{A-12}]$$

For an example, an automobile engine converts chemical energy in gasoline into kinetic energy of the automobile. The amount of energy converted into kinetic energy per unit time is the power of the engine. The SI unit of power is the watt (W): 1 W is defined as 1 joule per second.

When electrical current flows between two points of different potential, the potential energy of each charge changes. The change in potential energy per unit charge is the potential difference (V), and the amount of charge flowing per unit time is the current (i). Therefore, the electrical power (P) is equal to

$$P = iV \quad [\text{A-13}]$$

where *i* is the current and *V* is the potential difference. Because 1 V = 1 J/C and 1 amp = 1 C/s, 1 W = 1 A·V:

$$(1 \text{ A})(1 \text{ V}) = (1 \text{ C/s})(1 \text{ J/C}) = 1 \text{ J/s} = 1 \text{ W}$$

For example, if a potential difference of 12 V applied to a heating coil produces a current of 1 amp, the rate of heat production is

$$P = iV = (1 \text{ A})(12 \text{ V}) = 12 \text{ W}$$

Direct and Alternating Current

Electrical power is normally supplied to equipment as either *direct current* (DC) or *alternating current* (AC). DC power is provided by two wires connecting the equipment to a power source. The power source maintains a constant potential difference between the two wires (Fig. A-5A). Many electronic circuits require DC, and chemical batteries produce DC.

AC power is usually provided as *single-phase* or *three-phase* AC. Single-phase AC is supplied uses two wires from the power source. The power source produces a potential difference between the two wires that varies sinusoidally with time (see Fig. A-5B). Single-phase AC is specified by its amplitude (the maximal voltage) and its frequency (the number of cycles per unit time). Normal household and building power is single-phase 117-V AC. Its actual peak amplitude is 1.4×117 V, or 165 V.

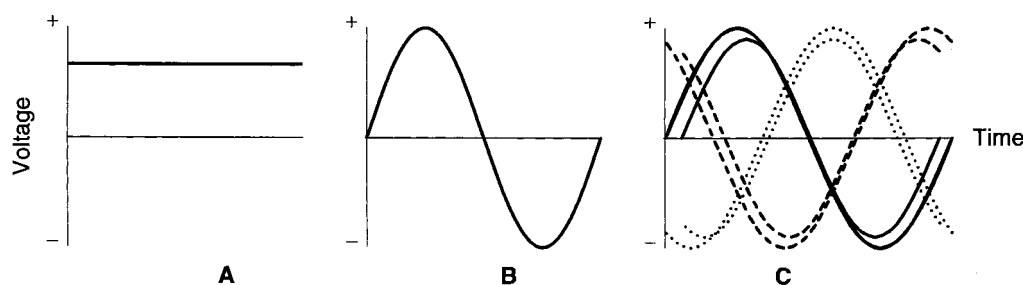


FIGURE A-5. **A:** Direct current. **B:** Single-phase alternating current. **C:** Three-phase alternating current.

AC power is usually supplied to buildings and to heavy-duty electrical equipment as *three-phase* AC. It uses three or four wires between the power source and the equipment. When present, the fourth wire, called the *neutral wire*, is maintained at ground potential (the potential of the earth). The power source provides a sinusoidally varying potential on each of the three wires with respect to the neutral wire. Voltages on these three wires have the same amplitudes and frequencies; however, each is a third of a cycle *out of phase* with respect to the other two, as shown in Fig. A-5C. Each cycle is assigned a total of 360 degrees, and therefore each cycle in a three-phase circuit is 120 degrees out of phase with the others.

AC is produced by electric generators and used by electric motors. The major advantage of AC over DC is that its amplitude may be easily increased or decreased using devices called *transformers*, as described later. AC is converted to DC for many electronic circuits.

Ground potential is the electrical potential of the earth. One of the two wires supplying single-phase AC is maintained at ground potential, as is the neutral wire used in supplying three-phase AC. These wires are maintained at ground potential by connecting them to a metal spike driven deep into the earth.

Conductors, Insulators, and Semiconductors

When a potential difference is applied across a conductor, it causes an electrical current to flow. In general, increasing the potential difference increases the current. The quantity *resistance* is defined as the ratio of the applied voltage to the resultant current:

$$R = V/i \quad [A-14]$$

The SI unit of resistance is the ohm (Ω): 1 Ω is defined as 1 volt per ampere.

Current is defined as the flow of positive charge. In solid matter, however, current is caused by the movement of negative charges (electrons) only. Based on the amount of current generated by an applied potential difference, solids may be roughly classified as conductors, semiconductors, or insulators. Metals, especially silver, copper, and aluminum, have very little resistance and are called conductors. Some materials, such as glass, plastics, and fused quartz, have very large resistances and are called insulators. Other materials, such as silicon and germanium, have intermediate resistances and are called *semiconductors*.

The band theory of solids explains the conduction properties of solids. The outer-shell electrons in solids exist in discrete energy bands. These bands are separated by gaps; electrons cannot possess energies within the gaps. For an electron to be mobile, there must be a nearby vacant position in the same energy band into which it can move. In conductors, the highest energy band occupied by electrons is only partially filled, so the electrons in it are mobile. In insulators and semiconductors, the highest occupied band is completely filled and electrons are not readily mobile.

The difference between insulators and semiconductors is the width of the forbidden gap between the highest occupied band and the next higher band; in semiconductors, the gap is about 1 eV, whereas in insulators it is typically greater than 5 eV. At room temperature, thermal energy temporarily promotes a small fraction of the electrons from the valence band into the next higher band. The promoted electrons are mobile. Their promotion also leaves behind vacancies in the valence band, called *holes*, into which other valence band electrons can move. Because the band gap of semiconductors is much smaller than that of insulators, a much larger number of electrons exists in the higher band of semiconductors and the resistance of semiconductors is therefore less than that of insulators. Reducing the temperature of insulators and semiconductors increases their resistance by reducing the thermal energy available for promoting electrons from the valence band.

For certain materials, including most metallic conductors, the resistance is not greatly affected by the applied voltage. In these materials, the applied voltage and resultant current are nearly proportional:

$$V = iR \quad [A-15]$$

This equation is called Ohm's Law.

Magnetism

Magnetic Forces and Fields

Magnetic fields can be caused by moving electrical charges. This charge motion may be over long distances, such as the movement of electrons through a wire, or it may be restricted to the vicinity of a molecule or atom, such as the motion of an unpaired electron in a valence shell. The basic features of magnetic fields can be illustrated by the bar magnet. First, the bar magnet, made of an iron ferromagnetic material, has a unique atomic electron orbital packing scheme that gives rise to an intense magnetic field due to the constructive addition of magnetic fields arising from *unpaired* spinning electrons in different atomic orbital shells. The resultant "magnet" has two poles and experiences a force when placed in the vicinity of another bar magnet or external magnetic field. By convention, the pole that points north under the influence of the earth's magnetic field is termed the *north pole* and the other is called the *south pole*. Experiments demonstrate that like poles of two magnets repel each other and unlike poles attract. When a magnet is broken in two, each piece becomes a new magnet, with a north and south pole. This phenomenon continues on to the atomic level. Therefore, the simplest magnetic structure is the magnetic *dipole*. In contrast, the simplest electrical structure, the isolated point charge, is unipolar.

Magnetic fields may be visualized as lines of magnetic force. Because magnetic poles exist in pairs, the magnetic field lines have no beginning or end and in fact circle upon themselves. Figure A-6A shows the magnetic field surrounding a bar

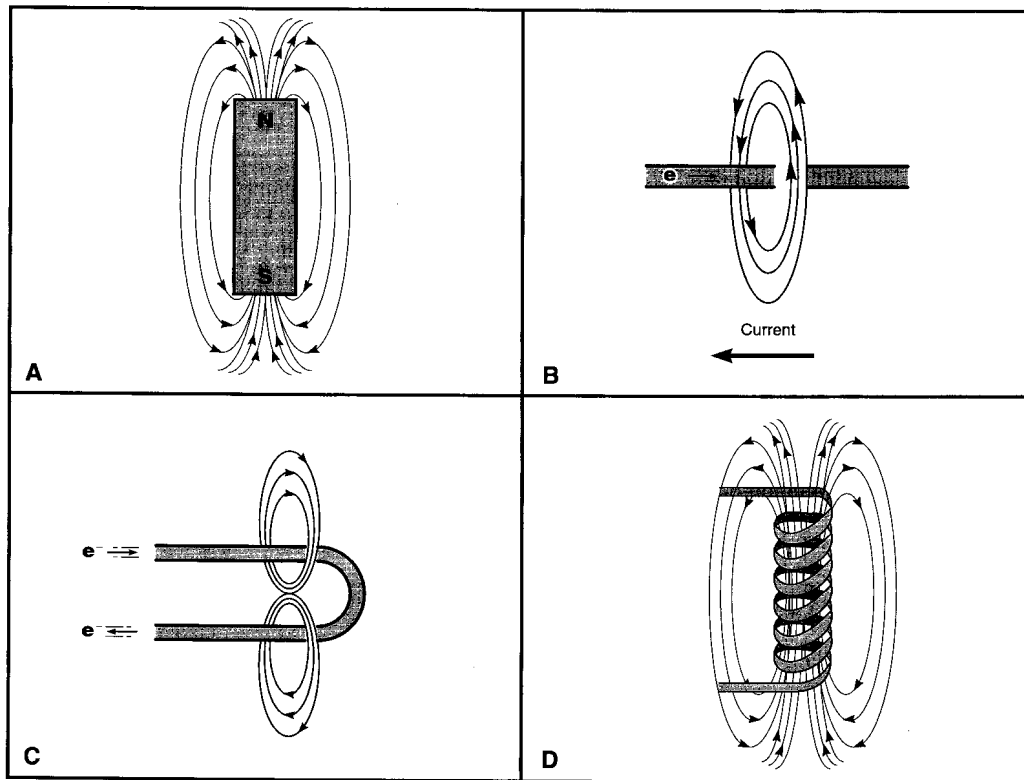


FIGURE A-6. Magnetic field descriptions: surrounding a bar magnet (**A**), around a wire (**B**), about a wire loop (**C**), and about a coiled wire (**D**).

magnet. A current carrying wire also produces a magnetic field that circles the wire, as shown in Fig. A-6B. Increasing the current flowing through the wire increases the magnetic field strength. The *right hand rule* allows the determination of the magnetic field direction by grasping the wire with the thumb pointed in the direction of the current (opposite the electron flow). The fingers will then circle the wire in the direction of the magnetic field. When a current carrying wire is curved in a loop, the resultant concentric lines of magnetic force overlap and augment the total local magnetic field strength inside the loop, as shown in Fig. A-6C. A coiled wire, called a *solenoid*, results in even more augmentation of the magnetic lines of force within the coil, as shown in Fig. A-6D. The field strength depends on the number of turns in the coil over a fixed distance. An extreme example of a solenoid is found in most magnetic resonance scanners; in fact, the patient resides inside the solenoid during the scan. A further enhancement of magnetic field strength can be obtained by applying a greater current through the wire or by placing a ferromagnetic material such as iron inside the solenoid. The iron core in this instance confines and augments the magnetic field. The solenoid is similar to the bar magnet discussed previously; however, the magnetic field strength may be changed by varying the current. If the current remains fixed (e.g., DC) the magnetic field also remains fixed; if the current varies (e.g., AC), the magnetic field varies. Solenoids are also called *electromagnets*.

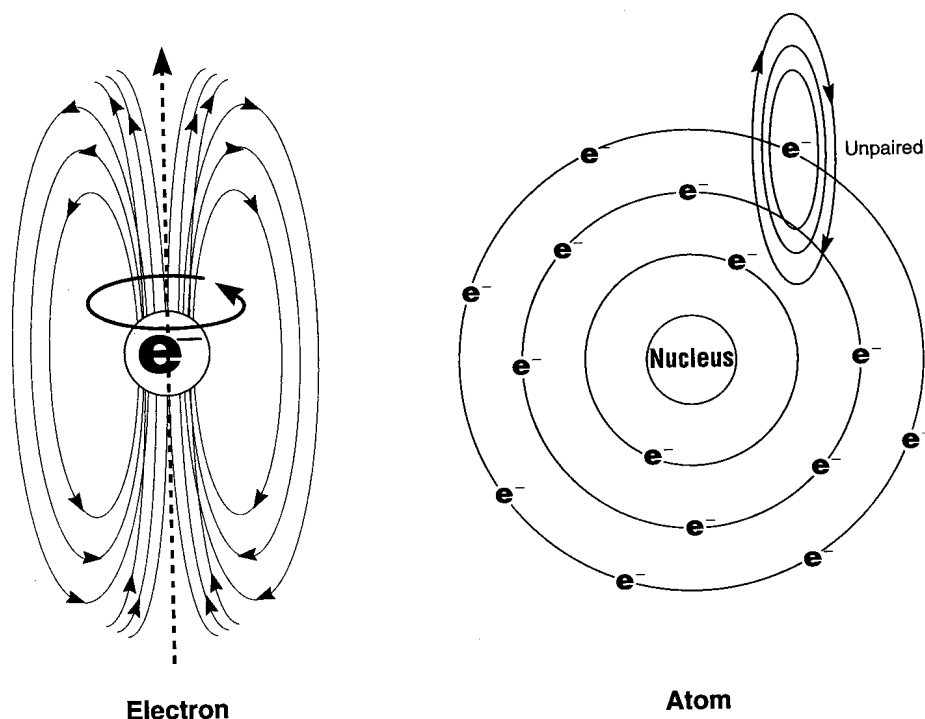


FIGURE A-7. A spinning charge has an associated magnetic field, such as the electron (left). An unpaired electron in an atomic orbital produces a magnetic field (right). Each produces magnetic dipoles that will interact with an external magnetic field.

Current loops behave as magnetic dipoles. Like the simple bar magnet, they tend to align with an external magnetic field, as shown in Fig. A-7. An electron in an atomic orbital is a current loop about the nucleus. A spinning charge such as the proton also may be thought of as a small current loop. These current loops act as magnetic dipoles and tend to align with external magnetic fields, a tendency that gives rise to the macroscopic magnetic properties of materials (discussed further in Chapter 14).

The SI unit of magnetic field strength is the *tesla* (T). An older unit of magnetic field strength is the *gauss* (G): 1 T is equal to 10^4 G. By way of comparison, the earth's magnetic field is approximately 0.5 to 1.0 G, whereas magnetic fields used for magnetic resonance imaging (MRI) typically range from 0.5 to 2 T.

Magnetic Forces on Moving Charged Particles

A magnetic field exerts a force on a moving charge, provided that the charge crosses the lines of magnetic field strength. The magnitude of this force is proportional to (a) the charge; (b) the speed of the charge; (c) the magnetic field strength, designated B ; and (d) the direction of charge travel with respect to the direction of the magnetic field. The direction of the force on the moving charge can be determined through the use of the right hand rule. It is perpendicular to both the direction of the charge's velocity and the lines of magnetic field strength. Because an electrical current consists of moving charges, a current-carrying wire in a magnetic field will experience a force. This principle is used in devices such as electric motors.

Electromagnetic Induction

In 1831, Michael Faraday discovered that a *moving* magnet induces an electrical current in a nearby conductor. His observations with magnets and conducting wires led to the findings listed here. The major ideas are illustrated in Fig. A-8.

1. A *changing* magnetic field induces a voltage (electrical potential difference) in a nearby conducting wire and causes a current to flow. An identical voltage is induced whether the wire moves with respect to the magnetic field position or

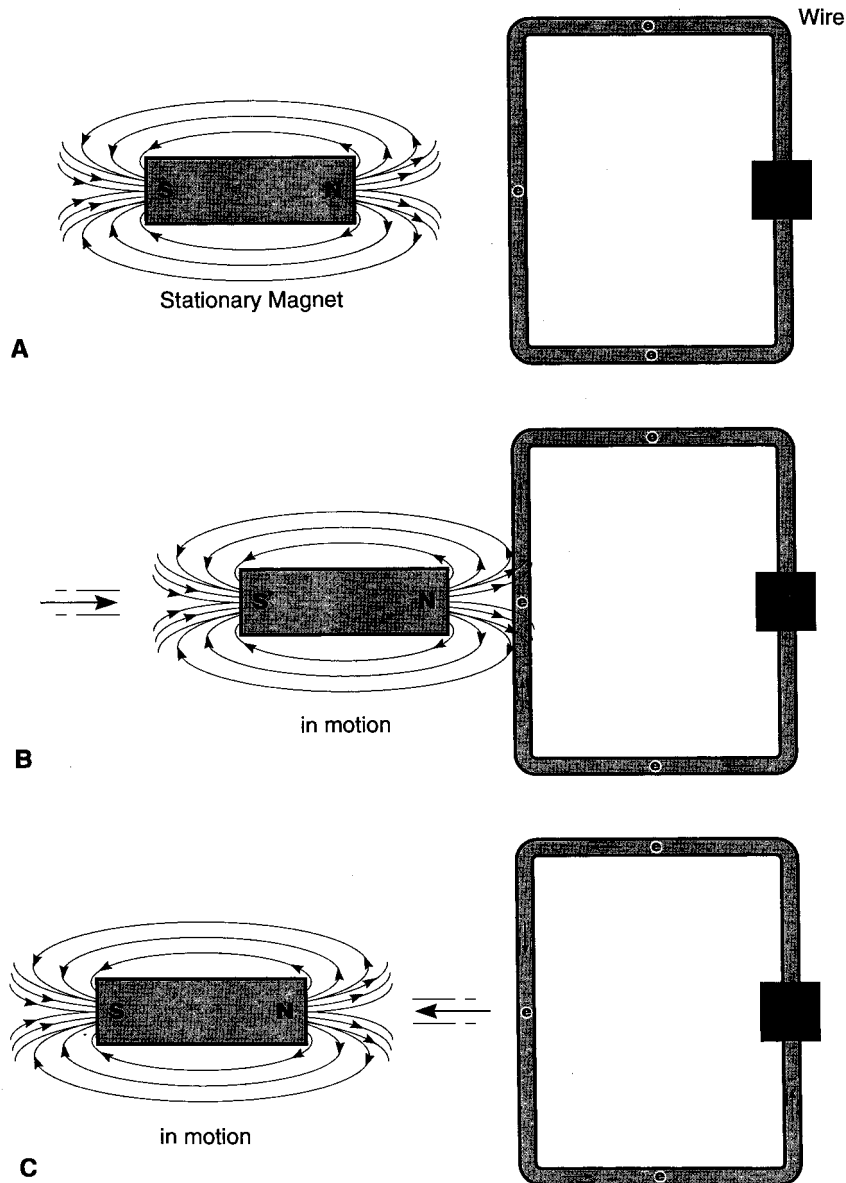


FIGURE A-8. Faraday experiments: changing magnetic fields and induction effects.

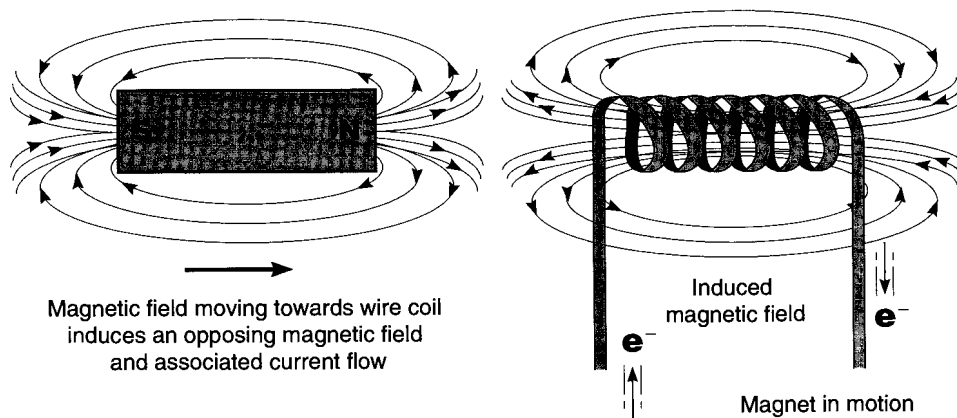


FIGURE A-9. Lenz's Law, demonstrating mutual induction between a moving magnetic field and a coiled wire conductor.

the magnetic field moves with respect to the wire position. A *stationary* magnetic field *does not* induce a voltage in a stationary wire.

2. A stronger magnetic field results in a stronger induced voltage. The voltage is proportional to the number of magnetic field lines cutting across the wire conductor per unit time. If the relative speed of the wire or the magnet is increased with respect to the other, the induced voltage will be larger because an increased number of magnetic field lines will be cutting across the wire conductor per unit time.
3. A 90-degree angle of the wire conductor relative to the magnetic field will provide the greatest number of lines to cross per unit distance, resulting in a higher induced voltage than for other angles.
4. When a solenoid (wire coil) is placed in a magnetic field, the magnetic field lines cut by each turn of the coil are additive, causing the resultant induced voltage to be directly proportional to the number of turns of the coil.

The direction of current flow caused by an induced voltage is described by Lenz's Law (Fig. A-9: *An induced current flows in a direction such that its associated magnetic field opposes the magnetic field that induced it.* Lenz's Law is an important concept that describes self-induction and mutual induction.

Self-Induction

A time-varying current in a coil wire produces a magnetic field that varies in the same manner. By Lenz's Law, the varying magnetic field induces a potential difference across the coil that opposes the source voltage. Therefore, a rising and falling source voltage (AC) creates an *opposing* falling and rising induced voltage in the coil. This phenomenon, called *self-induction*, is used in autotransformers that provide a variable voltage, as discussed in Chapter 5.

Mutual Induction

A *primary* wire coil carrying AC produces a time-varying magnetic field. When a *secondary* wire coil is located under the influence of this magnetic field, a time-vary-

ing potential difference across the secondary coil is similarly induced, as shown in Fig. A-9. The amplitude of the induced voltage, V_s , can be determined from the following equation, known as the law of transformers:

$$\left(\frac{V_p}{N_p}\right) = \left(\frac{V_s}{N_s}\right) \quad [\text{A-16}]$$

where V_p is the voltage amplitude applied to the primary coil, V_s is the voltage amplitude induced in the secondary coil, N_p is the number of turns in the primary coil, and N_s is the number of turns in the secondary coil. With a predetermined number of “turns” on the primary and secondary coils, this *mutual inductance* property can increase or decrease the voltage in electrical circuits. The devices that provide this change of voltage (current changes in the opposite direction) are called *transformers*. Further explanation of their use in x-ray generators is covered in Chapter 5.

Electric Generators and Motors

The electric generator utilizes the principles of electromagnetic induction to convert mechanical energy into electrical energy. It consists of a coiled wire mounted on a rotor between the poles of a strong magnet, as shown in Fig. A-10A. As an external mechanical energy source rotates the coil (e.g., hydroelectric turbine generator), the wires in the coil cut across the magnetic force lines, resulting in a sinusoidally varying potential difference, the polarity of which is determined by the wire approaching or receding from one pole of the magnet. The generator serves as a source of AC power.

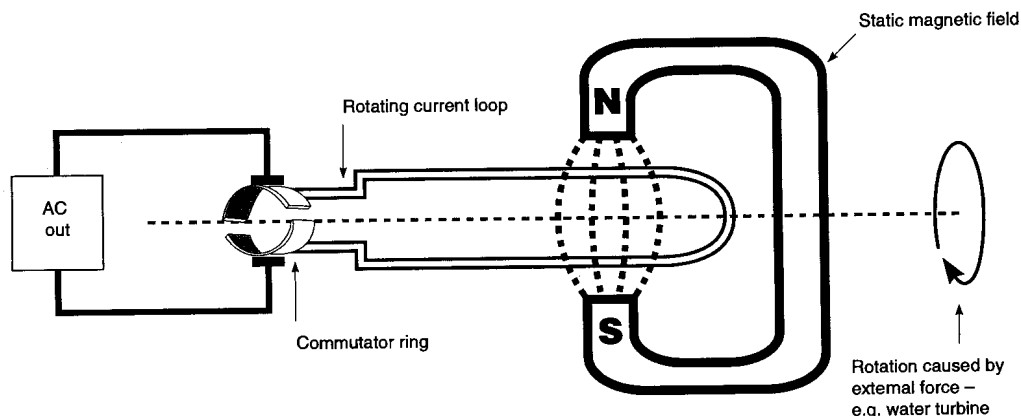
The electric motor converts electrical energy into mechanical energy. It consists of a coil of wires mounted on a freely rotating axis (rotor) between the poles of a fixed magnet, as shown in Fig. A-10B. When AC flows through the coil, an increasing and decreasing magnetic field is generated, the coil acting as a magnetic dipole. This dipole tends to align with the external magnetic field, causing the rotor to turn. As the dipole approaches alignment, however, the AC and thus the magnetic field of the rotor reverses polarity, causing another half-turn rotation to achieve alignment. The alternating polarity of the rotor’s magnetic field causes a continuous rotation of the coil wire mounted on the rotor as long as AC is applied.

Magnetic Properties of Matter

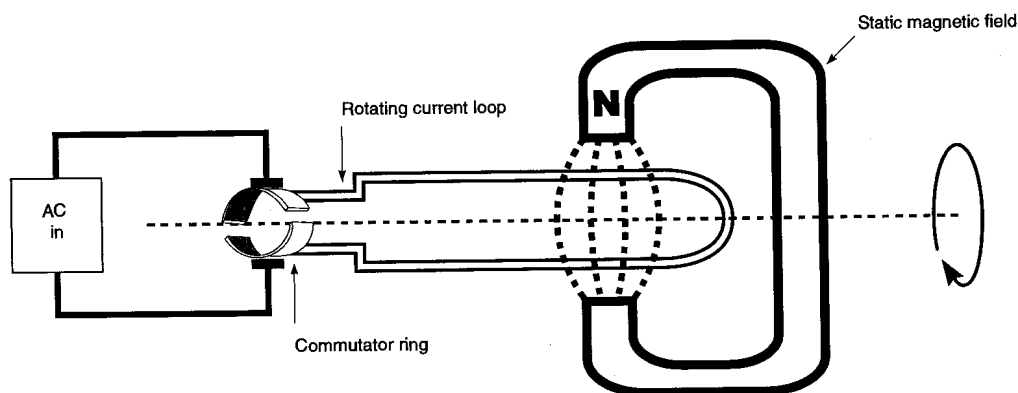
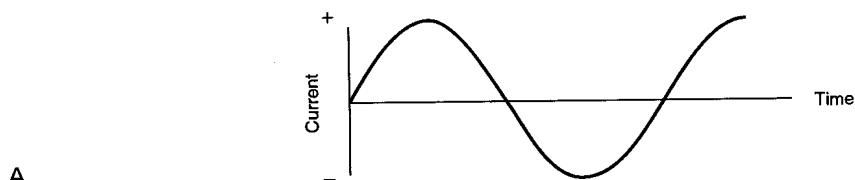
The magnetic characteristics of materials are determined by atomic and molecular structures related to the behavior of the associated electrons. Three categories of magnetic properties are defined: diamagnetic, paramagnetic, and ferromagnetic. Their properties arise from the association of moving charges (electrons) and magnetic fields.

Diamagnetic Materials

Individual electrons orbiting in atomic or molecular shells represent a current loop and give rise to a magnetic field. The various electron orbits do not fall in any preferred plane, so the superposition of the associated magnetic fields results in a net magnetic field that is too small to be measured. When these atoms or molecules are placed in a changing magnetic field, however, the electron motions are altered by the induced electrical motive force to form a reverse magnetic field *opposing* the applied magnetic field. *Diamagnetic materials* therefore cause a depletion of the applied magnetic field in the local micromagnetic environment.



The rotating current wire loop in the static magnetic field creates an induced current occurring with an alternating potential difference.



B Induced rotation of the current loop is caused by the variable magnetic field within current loop associated with the applied alternating current interacting with the static magnetic field.

FIGURE A-10. Electrical and mechanical energy conversion: electric generators **(A)** and electric motors **(B)**.

Paramagnetic Materials

Based on the previous discussion, it would seem likely that all elements and molecules would behave as diamagnetic agents. In fact, all have diamagnetic properties, but some have additional properties that overwhelm the diamagnetic response. Electrons in atomic or molecular orbitals orient their magnetic fields in such a way

as to cancel in pairs; however, in those materials that have an *odd* number of electrons, one electron is unpaired, exact cancellation cannot occur, and a magnetic field equivalent to one electron orbital results. Depending on orbital structures and electron filling characteristics, *fractional* unpaired electron spins can also occur, resulting in variations in the magnetic field strength. When placed in an external magnetic field, the magnetic field of the material caused by the unpaired electron aligns with the applied field. *Paramagnetic materials*, having unpaired electrons, locally *augment* the micromagnetic environment. The magnetic overall properties of the atom result from these paramagnetic effects as well as the diamagnetic effects discussed earlier, with the paramagnetic characteristics predominating.

Ferromagnetic Materials

Iron, nickel, and cobalt can possess intrinsic magnetic fields and will react strongly in an applied magnetic field. These *ferromagnetic materials* are transition elements that have an unorthodox atomic orbital structure: Electrons will fill the outer orbital shells before the inner orbitals are completely filled. The usual spin cancellation of the electrons does not occur, resulting in an unusually high atomic magnetic moment. While they are in a random atomic (elemental) or molecular (compound) arrangement, cancellation of the dipoles occurs and no intrinsic magnetic field is manifested; however, when the individual magnetic dipoles are nonrandomly aligned by an external force (e.g., a strong electrical field), a constructive enhancement of the individual atomic magnetic moments gives rise to an intrinsic magnetic field that reacts strongly in an applied magnetic field. Permanent magnets are examples of a permanent *nonrandom* arrangement of local “magnetic domains” of the transition elements. Ferromagnetic characteristics dominate over paramagnetic and diamagnetic interactions.

Appendix B

PHYSICAL CONSTANTS, PREFIXES, GEOMETRY, CONVERSION FACTORS, AND RADIOLOGIC DATA

**TABLE B-1. PHYSICAL CONSTANTS, PREFIXES, AND
GEOMETRY**

A: Physical Constants

Acceleration of gravity	9.80665 m/s ²
Gyromagnetic ratio of proton	42.5764 MHz/tesla
Avogadro's number	6.0221367×10^{23} particles/mole
Planck's constant	6.62607×10^{-34} J-s
Planck's constant	4.13570×10^{-15} eV-s
Charge of electron	1.602177×10^{-19} coulomb
Mass of electron	9.109384×10^{-31} kg
Mass of proton	1.672623×10^{-27} kg
Mass of neutron	1.674929×10^{-27} kg
Atomic mass unit	1.660570×10^{-27} kg
Molar volume	22.4138 L/mole
Speed of light (c)	2.99792458×10^8 m/s

B: Prefixes

yotta (Y)	10 ²⁴
zetta (Z)	10 ²¹
exa (E)	10 ¹⁸
peta (P)	10 ¹⁵
tera (T)	10 ¹²
giga (G)	10 ⁹
mega (M)	10 ⁶
kilo (k)	10 ³
hecto (h)	10 ²
deca (da)	10 ¹
deci (d)	10 ⁻¹
centi (c)	10 ⁻²
milli (m)	10 ⁻³
micro (μ)	10 ⁻⁶
nano (n)	10 ⁻⁹
pico (p)	10 ⁻¹²
femto (f)	10 ⁻¹⁵
atto (a)	10 ⁻¹⁸
zepto (z)	10 ⁻²¹
yocto (y)	10 ⁻²⁴

(continued)

TABLE B-1 (continued).**C: Geometry**

Area of circle	πr^2
Circumference of circle	$2\pi r$
Surface area of sphere	$4\pi r^2$
Volume of sphere	$(4/3)\pi r^3$
Area of triangle	$(1/2) \text{ height} \times \text{base}$
1 radian	$360^\circ/2\pi = 57.2958 \text{ degrees}$

TABLE B-2. CONVERSION FACTORS**Energy**

1 electron volt (eV)	$1.6021 \times 10^{-19} \text{ joule}$
1 calorie	4.187 joule

Power

1 kilowatt (kW)	joule/s
-----------------	---------

Current

1 ampere	1 coulomb/s
1 ampere	$6.281 \times 10^{18} \text{ electrons/s}$

Volume

1 U.S. gallon	3.7853 L
1 liter (L)	1000 cm ³

Temperature

°C (Celsius)	$5/9 \times (^\circ\text{F} - 32)$
°F (Fahrenheit)	$9/5 \times ^\circ\text{C} + 32$
°K (Kelvin)	$^\circ\text{C} + 273.16$

Mass

1 pound	2.205 kg
---------	----------

Length

1 inch	25.4 mm
1 inch	2.54 cm

Radiologic Units

1 roentgen (R)	$2.58 \times 10^{-4} \text{ coulomb/kg}$
1 roentgen (R)	8.708 mGy air kerma (@30 kVp)
1 roentgen (R)	8.767 mGy air kerma (@60 kVp)
1 roentgen (R)	8.883 mGy air kerma (@100 kVp)
1 mGy air kerma	0.114 roentgen (@60 kVp)
1 gray (Gy)	100 rad
1 sievert (Sv)	100 rem
1 curie (Ci)	$3.7 \times 10^{10} \text{ becquerel (Bq)}$

TABLE B-3. RADIOLOGIC DATA FOR ELEMENTS 1 THROUGH 100

Z	Sym	Element	Density g/cm ³	At mass g/mole	K-edge keV	L-edge keV	K _{α1} keV	K _{α2} keV	K _{β1} keV	K _{β2} keV
1	H	Hydrogen	0.0001	1.008	0.01	0.00	0.01	0.01	0.01	0.01
2	He	Helium	0.0002	4.003	0.02	0.00	0.02	0.02	0.02	0.02
3	Li	Lithium	0.5340	6.939	0.06	0.00	0.06	0.06	0.06	0.06
4	Be	Beryllium	1.8480	9.012	0.12	0.00	0.12	0.12	0.12	0.12
5	B	Boron	2.3700	10.811	0.20	0.01	0.19	0.19	0.20	0.20
6	C	Carbon	2.2650	12.011	0.29	0.01	0.28	0.28	0.29	0.29
7	N	Nitrogen	0.0013	14.007	0.41	0.02	0.39	0.39	0.41	0.41
8	O	Oxygen	0.0014	15.999	0.54	0.02	0.52	0.52	0.54	0.54
9	F	Fluorine	0.0017	18.998	0.69	0.02	0.67	0.67	0.69	0.69
10	Ne	Neon	0.0009	20.179	0.86	0.03	0.84	0.84	0.86	0.86
11	Na	Sodium	0.9710	22.990	1.06	0.05	1.03	1.03	1.05	1.06
12	Mg	Magnesium	1.7400	24.312	1.29	0.07	1.24	1.24	1.28	1.29
13	Al	Aluminum	2.6990	26.982	1.55	0.09	1.47	1.47	1.54	1.55
14	Si	Silicon	2.3300	28.086	1.83	0.12	1.72	1.72	1.82	1.83
15	P	Phosphorus	2.2000	30.974	2.13	0.15	1.99	1.99	2.12	2.13
16	S	Sulphur	2.0700	32.064	2.46	0.19	2.29	2.28	2.45	2.46
17	Cl	Chlorine	0.0032	35.453	2.81	0.23	2.60	2.60	2.79	2.81
18	Ar	Argon	0.0018	39.948	3.18	0.27	2.93	2.93	3.16	3.18
19	K	Potassium	0.8620	39.102	3.58	0.32	3.28	3.28	3.56	3.57
20	Ca	Calcium	1.5500	40.080	4.01	0.38	3.66	3.66	3.98	4.00
21	Sc	Scandium	2.9890	44.958	4.47	0.44	4.06	4.05	4.46	4.46
22	Ti	Titanium	4.5400	47.900	4.94	0.50	4.48	4.47	4.93	4.93
23	V	Vanadium	6.1100	50.942	5.44	0.56	4.91	4.91	5.43	5.43
24	Cr	Chromium	7.1800	51.996	5.96	0.62	5.38	5.37	5.95	5.95
25	Mn	Manganese	7.4400	54.938	6.51	0.69	5.86	5.85	6.50	6.50
26	Fe	Iron	7.8740	55.847	7.08	0.77	6.36	6.35	7.07	7.07
27	Co	Cobalt	8.9000	58.933	7.68	0.84	6.89	6.87	7.67	7.67
28	Ni	Nickel	8.9020	58.710	8.30	0.92	7.44	7.42	8.29	8.29
29	Cu	Copper	8.9600	63.540	8.94	0.99	8.01	7.98	8.93	8.93
30	Zn	Zinc	7.1330	65.370	9.62	1.09	8.60	8.57	9.61	9.61
31	Ga	Gallium	5.9040	69.720	10.33	1.19	9.21	9.18	10.30	10.32
32	Ge	Germanium	5.3230	72.590	11.07	1.29	9.84	9.81	11.03	11.06
33	As	Arsenic	5.7300	74.922	11.83	1.40	10.50	10.46	11.78	11.82
34	Se	Selenium	4.5000	78.960	12.62	1.52	11.18	11.14	12.56	12.61
35	Br	Bromine	0.0076	79.909	13.44	1.64	11.88	11.83	13.36	13.42
36	Kr	Krypton	0.0037	83.800	14.28	1.77	12.61	12.55	14.19	14.27
37	Rb	Rubidium	1.5320	85.470	15.16	1.91	13.35	13.29	15.04	15.14
38	Sr	Strontium	2.5400	87.620	16.07	2.05	14.12	14.05	15.92	16.04
39	Y	Yttrium	4.4690	88.905	17.00	2.20	14.92	14.84	16.83	16.99
40	Zr	Zirconium	6.5060	91.220	17.96	2.35	15.74	15.65	17.77	17.95
41	Nb	Niobium	8.5700	92.906	18.95	2.50	16.58	16.48	18.74	18.94
42	Mo	Molybdenum	10.2200	95.940	19.97	2.67	17.44	17.34	19.73	19.95
43	Tc	Technetium	11.5000	99.000	21.01	2.83	18.33	18.21	20.75	21.00
44	Ru	Ruthenium	12.4100	101.700	22.09	3.00	19.25	19.11	21.80	22.08
45	Rh	Rhodium	12.4100	102.905	23.19	3.18	20.19	20.04	22.87	23.18
46	Pd	Palladium	12.0200	106.400	24.32	3.36	21.15	20.99	23.98	24.31
47	Ag	Silver	10.5000	107.870	25.49	3.55	22.14	21.96	25.11	25.48
48	Cd	Cadmium	8.6500	112.400	26.69	3.75	23.15	22.96	26.27	26.67
49	In	Indium	7.3100	114.820	27.92	3.96	24.19	23.98	27.47	27.90
50	Sn	Tin	7.3100	118.690	29.18	4.18	25.26	25.02	28.69	29.15
51	Sb	Antimony	6.6910	121.750	30.48	4.40	26.35	26.09	29.94	30.44
52	Te	Tellurium	6.2400	127.600	31.81	4.62	27.47	27.19	31.22	31.76
53	I	Iodine	4.9300	126.904	33.16	4.86	28.61	28.31	32.53	33.11

(continued)

TABLE B-3 (continued).

Z	Sym	Element	Density g/cm ³	At mass g/mole	K-edge keV	L-edge keV	K _{α1} keV	K _{α2} keV	K _{β1} keV	K _{β2} keV
54	Xe	Xenon	0.0059	131.300	34.56	5.10	29.78	29.45	33.88	34.49
55	Cs	Cesium	1.8730	132.905	35.99	5.35	30.98	30.62	35.25	35.90
56	Ba	Barium	3.5000	137.340	37.45	5.61	32.21	31.82	36.66	37.35
57	La	Lanthanum	6.1540	138.910	38.94	5.88	33.46	33.04	38.10	38.83
58	Ce	Cerium	6.6570	140.120	40.46	6.13	34.74	34.29	39.57	40.45
59	Pr	Praseodymium	6.7100	140.907	42.01	6.40	36.06	35.57	41.07	42.00
60	Nd	Neodymium	6.9000	144.240	43.60	6.67	37.40	36.87	42.61	43.59
61	Pm	Promethium	7.2200	145.000	45.22	6.96	38.77	38.20	44.18	45.21
62	Sm	Samarium	7.4600	150.350	46.88	7.24	40.17	39.56	45.79	46.87
63	Eu	Europium	5.2430	151.960	48.57	7.54	41.60	40.94	47.43	48.56
64	Gd	Gadolinium	7.9000	157.250	50.30	7.85	43.06	42.36	49.10	50.29
65	Tb	Terbium	8.2290	158.924	52.06	8.15	44.56	43.80	50.82	52.05
66	Dy	Dysprosium	8.5500	162.500	53.86	8.46	46.08	45.27	52.56	53.85
67	Ho	Holmium	8.7950	164.930	55.70	8.78	47.64	46.77	54.34	55.69
68	Er	Erbium	9.0660	167.260	57.57	9.11	49.23	48.30	56.16	57.56
69	Tm	Thulium	9.3210	168.934	59.49	9.45	50.85	49.86	58.02	59.48
70	Yb	Ytterbium	6.7300	173.040	61.44	9.79	52.51	51.45	59.91	61.43
71	Lu	Lutecium	9.8490	174.970	63.44	10.15	54.20	53.07	61.84	63.43
72	Hf	Hafnium	13.3100	178.490	65.48	10.52	55.92	54.72	63.81	65.46
73	Ta	Tantalum	16.6540	180.948	67.57	10.90	57.69	56.40	65.82	67.53
74	W	Tungsten	19.3000	183.850	69.69	11.29	59.48	58.12	67.87	69.65
75	Re	Rhenium	21.0200	186.200	71.86	11.68	61.32	59.86	69.96	71.80
76	Os	Osmium	22.5700	190.200	74.07	12.08	63.19	61.64	72.10	74.00
77	Ir	Iridium	22.4200	192.200	76.31	12.49	65.10	63.46	74.27	76.25
78	Pt	Platinum	21.4500	195.090	78.61	12.91	67.05	65.30	76.49	78.53
79	Au	Gold	19.3200	196.967	80.96	13.34	69.04	67.18	78.75	80.87
80	Hg	Mercury	13.5460	200.500	83.35	13.79	71.07	69.10	81.05	83.24
81	Tl	Thallium	11.7200	204.370	85.80	14.25	73.14	71.05	83.40	85.67
82	Pb	Lead	11.3500	207.190	88.29	14.71	75.25	73.04	85.80	88.14
83	Bi	Bismuth	9.7470	208.980	90.83	15.19	77.41	75.06	88.25	90.66
84	Po	Polonium	9.3200	209.000	93.42	15.68	79.61	77.13	90.74	93.24
85	At	Astatine	9.0000	210.000	96.07	16.18	81.85	79.23	93.28	95.86
86	Rn	Radon	0.0097	222.000	98.76	16.69	84.14	81.37	95.88	98.53
87	Fr	Francium	5.0000	223.000	101.52	17.21	86.49	83.55	98.52	101.26
88	Ra	Radium	5.0000	226.000	104.32	17.75	88.86	85.76	101.21	104.03
89	Ac	Actinium	10.0700	227.000	107.19	18.30	91.31	88.03	103.97	106.87
90	Th	Thorium	11.7200	232.038	110.11	18.86	93.79	90.33	106.77	109.76
91	Pa	Protactinium	15.3700	233.000	113.08	19.42	96.34	92.68	109.64	112.71
92	U	Uranium	18.9500	238.051	116.11	20.00	98.93	95.07	112.56	115.72
93	Np	Neptunium	20.2500	237.000	119.20	20.59	101.57	97.50	115.53	118.78
94	Pu	Plutonium	19.8400	239.052	122.35	21.19	104.28	99.98	118.57	121.92
95	Am	Americium	13.6700	242.000	125.57	21.81	107.04	102.51	121.68	125.11
96	Cm	Curium	13.5100	247.000	128.86	22.44	109.87	105.08	124.84	128.37
97	Bk	Berkelium	14.0000	247.000	132.21	23.09	112.75	107.71	128.08	131.70
98	Cf	Californium	14.0000	251.000	135.63	23.74	115.70	110.39	131.38	135.10
99	Es	Einsteinium	14.0000	252.000	139.11	24.41	118.70	113.10	134.74	138.55
100	Fm	Fermium	14.0000	257.000	142.68	25.10	121.79	115.89	138.19	142.10

Appendix C

MASS ATTENUATION COEFFICIENTS AND SPECTRA DATA TABLES

C.1 MASS ATTENUATION COEFFICIENTS FOR SELECTED ELEMENTS

**TABLE C-1. MASS ATTENUATION COEFFICIENTS IN CM²/G (DENSITY (ρ)
IN G/CM³)**

Energy (keV)	Aluminum Z = 13 ρ = 2.699	Calcium Z = 20 ρ = 1.55	Copper Z = 29 ρ = 8.96	Molybdenum Z = 42 ρ = 10.22	Rhodium Z = 45 ρ = 12.41	Iodine Z = 53 ρ = 4.93	Tungsten Z = 74 ρ = 19.3	Lead Z = 82 ρ = 11.35
2	2.262E + 3	7.952E + 2	2.145E + 3	9.559E + 2	1.212E + 3	1.994E + 3	3.914E + 3	1.276E + 3
3	7.830E + 2	2.712E + 2	7.466E + 2	2.009E + 3	4.422E + 2	7.491E + 2	1.922E + 3	1.958E + 3
4	3.581E + 2	1.199E + 2	3.459E + 2	9.651E + 2	1.168E + 3	3.560E + 2	9.545E + 2	1.248E + 3
5	1.924E + 2	5.973E + 2	1.890E + 2	5.417E + 2	6.548E + 2	8.299E + 2	5.514E + 2	7.283E + 2
6	1.149E + 2	3.737E + 2	1.151E + 2	3.377E + 2	4.110E + 2	6.099E + 2	3.552E + 2	4.729E + 2
7	7.365E + 1	2.457E + 2	7.480E + 1	2.229E + 2	2.716E + 2	4.126E + 2	2.356E + 2	3.157E + 2
8	5.038E + 1	1.717E + 2	5.180E + 1	1.562E + 2	1.905E + 2	2.886E + 2	1.686E + 2	2.271E + 2
9	3.555E + 1	1.251E + 2	2.758E + 2	1.135E + 2	1.394E + 2	2.133E + 2	1.235E + 2	1.680E + 2
10	2.605E + 1	9.302E + 1	2.141E + 2	8.549E + 1	1.049E + 2	1.623E + 2	9.333E + 1	1.287E + 2
12	1.533E + 1	5.601E + 1	1.355E + 2	5.223E + 1	6.438E + 1	1.003E + 2	2.101E + 2	8.152E + 1
14	9.731E + 0	3.620E + 1	8.914E + 1	3.399E + 1	4.211E + 1	6.609E + 1	1.646E + 2	1.326E + 2
16	6.576E + 0	2.477E + 1	6.204E + 1	2.369E + 1	2.917E + 1	4.612E + 1	1.173E + 2	1.506E + 2
18	4.647E + 0	1.763E + 1	4.486E + 1	1.722E + 1	2.105E + 1	3.351E + 1	8.632E + 1	1.104E + 2
20	3.423E + 0	1.302E + 1	3.359E + 1	7.894E + 1	1.573E + 1	2.521E + 1	6.560E + 1	8.606E + 1
25	1.830E + 0	6.884E + 0	1.822E + 1	4.779E + 1	5.286E + 1	1.372E + 1	3.657E + 1	4.928E + 1
30	1.131E + 0	4.067E + 0	1.090E + 1	2.821E + 1	3.321E + 1	8.382E + 0	2.284E + 1	3.062E + 1
35	7.692E - 1	2.631E + 0	7.051E + 0	1.851E + 1	2.207E + 1	3.103E + 1	1.507E + 1	2.022E + 1
40	5.675E - 1	1.825E + 0	4.854E + 0	1.289E + 1	1.536E + 1	2.198E + 1	1.057E + 1	1.427E + 1
45	4.461E - 1	1.337E + 0	3.506E + 0	9.385E + 0	1.118E + 1	1.620E + 1	7.719E + 0	1.052E + 1
50	3.684E - 1	1.023E + 0	2.619E + 0	7.069E + 0	8.419E + 0	1.235E + 1	5.845E + 0	7.983E + 0
55	3.151E - 1	8.071E - 1	2.014E + 0	5.424E + 0	6.524E + 0	9.560E + 0	4.537E + 0	6.202E + 0
60	2.778E - 1	6.606E - 1	1.600E + 0	4.287E + 0	5.181E + 0	7.606E + 0	3.617E + 0	4.978E + 0
65	2.517E - 1	5.560E - 1	1.303E + 0	3.465E + 0	4.200E + 0	6.181E + 0	2.920E + 0	4.039E + 0
70	2.302E - 1	4.719E - 1	1.063E + 0	2.803E + 0	3.393E + 0	5.021E + 0	1.089E + 1	3.297E + 0
75	2.144E - 1	4.122E - 1	9.000E - 1	2.331E + 0	2.841E + 0	4.186E - 0	9.214E + 0	2.770E + 0
80	2.018E - 1	3.655E - 1	7.631E - 1	1.961E + 0	2.368E + 0	3.529E + 0	7.805E + 0	2.316E + 0
85	1.924E - 1	3.329E - 1	6.714E - 1	1.669E + 0	2.050E + 0	3.047E + 0	6.781E + 0	1.977E + 0

(continued)

TABLE C-1 (continued).

Energy (keV)	Aluminum Z = 13 $\rho = 2.699$	Calcium Z = 20 $\rho = 1.55$	Copper Z = 29 $\rho = 8.96$	Molybdenum Z = 42 $\rho = 10.22$	Rhodium Z = 45 $\rho = 12.41$	Iodine Z = 53 $\rho = 4.93$	Tungsten Z = 74 $\rho = 19.3$	Lead Z = 82 $\rho = 11.35$
90	1.832E - 1	3.002E - 1	5.796E - 1	1.441E + 0	1.735E + 0	2.572E + 0	5.780E + 0	7.200E + 0
95	1.769E - 1	2.789E - 1	5.199E - 1	1.270E + 0	1.527E + 0	2.258E + 0	5.102E + 0	6.376E + 0
100	1.705E - 1	2.577E - 1	4.604E - 1	1.102E + 0	1.322E + 0	1.946E + 0	4.424E + 0	5.540E + 0
110	1.607E - 1	2.275E - 1	3.782E - 1	8.665E - 1	1.035E + 0	1.517E + 0	3.475E + 0	4.357E + 0
120	1.533E - 1	2.062E - 1	3.234E - 1	7.095E - 1	8.431E - 1	1.219E + 0	2.834E + 0	3.553E + 0
130	1.474E - 1	1.900E - 1	2.806E - 1	5.878E - 1	6.954E - 1	1.004E + 0	2.294E + 0	2.899E + 0
140	1.421E - 1	1.773E - 1	2.467E - 1	4.911E - 1	5.774E - 1	8.264E - 1	1.893E + 0	2.383E + 0
150	1.378E - 1	1.674E - 1	2.226E - 1	4.230E - 1	4.943E - 1	7.004E - 1	1.594E + 0	2.016E + 0
160	1.339E - 1	1.593E - 1	2.032E - 1	3.695E - 1	4.291E - 1	6.015E - 1	1.356E + 0	1.715E + 0
170	1.306E - 1	1.528E - 1	1.891E - 1	3.314E - 1	3.824E - 1	5.307E - 1	1.185E + 0	1.500E + 0
180	1.275E - 1	1.466E - 1	1.751E - 1	2.932E - 1	3.362E - 1	4.606E - 1	1.017E + 0	1.289E + 0
190	1.247E - 1	1.420E - 1	1.655E - 1	2.680E - 1	3.056E - 1	4.142E - 1	9.034E - 1	1.146E + 0
200	1.219E - 1	1.374E - 1	1.560E - 1	2.429E - 1	2.750E - 1	3.677E - 1	7.904E - 1	1.003E + 0

C.2 MASS ATTENUATION COEFFICIENTS FOR SELECTED COMPOUNDS

TABLE C-2. MASS ATTENUATION COEFFICIENTS IN CM²/G (DENSITY (ρ) IN G/CM³)

Energy (keV)	Air $\rho = 0.001293$	Water $\rho = 1.00$	Plexiglas ^a $\rho = 1.19$	Muscle $\rho = 1.06$	Bone $\rho = 1.5$ to 3.0	Adipose ^b $\rho = 0.930$	50/50 ^b $\rho = 0.982$	Glandular ^b $\rho = 1.040$
2	6.172E + 2	2.880E + 3	2.111E + 3	2.481E + 3	2.367E + 3	2.764E + 3	1.604E + 3	1.614E + 3
3	1.954E + 2	1.219E + 3	7.964E + 2	9.202E + 2	1.487E + 3	1.314E + 3	6.571E + 2	6.420E + 2
4	8.240E + 1	5.899E + 2	3.796E + 2	4.369E + 2	7.276E + 2	6.468E + 2	1.105E + 3	9.481E + 2
5	4.235E + 1	3.347E + 2	5.204E + 2	2.928E + 2	4.136E + 2	4.458E + 2	7.829E + 2	6.636E + 2
6	2.466E + 1	2.124E + 2	6.437E + 2	4.965E + 2	2.643E + 2	2.850E + 2	5.311E + 2	4.482E + 2
7	1.542E + 1	1.393E + 2	4.328E + 2	3.393E + 2	1.738E + 2	1.885E + 2	3.569E + 2	3.007E + 2
8	1.029E + 1	3.436E + 2	3.044E + 2	2.369E + 2	1.232E + 2	1.341E + 2	2.583E + 2	2.172E + 2
9	7.233E + 0	2.961E + 2	2.245E + 2	1.731E + 2	9.005E + 1	9.810E + 1	1.922E + 2	1.613E + 2
10	5.296E + 0	2.274E + 2	1.708E + 2	1.318E + 2	1.459E + 2	7.386E + 1	1.479E + 2	1.240E + 2
12	3.162E + 0	1.431E + 2	1.056E + 2	8.158E + 1	1.384E + 2	1.427E + 2	9.400E + 1	7.867E + 1
14	2.015E + 0	9.470E + 1	6.966E + 1	8.538E + 1	9.288E + 1	1.107E + 2	6.241E + 1	5.215E + 1
16	1.402E + 0	6.644E + 1	4.870E + 1	6.006E + 1	6.697E + 1	7.870E + 1	4.416E + 1	3.686E + 1
18	1.038E + 0	4.859E + 1	3.540E + 1	4.394E + 1	6.926E + 1	5.782E + 1	7.731E + 1	6.361E + 1
20	8.077E - 1	3.683E + 1	2.665E + 1	3.324E + 1	5.260E + 1	4.389E + 1	8.205E + 1	6.731E + 1
25	5.088E - 1	2.024E + 1	1.452E + 1	1.818E + 1	2.928E + 1	2.443E + 1	5.477E + 1	4.488E + 1
30	3.753E - 1	1.250E + 1	8.872E + 0	1.113E + 1	1.821E + 1	1.523E + 1	3.441E + 1	2.820E + 1
35	3.071E - 1	8.223E + 0	1.823E + 1	9.205E + 0	1.199E + 1	1.006E + 1	2.307E + 1	1.890E + 1
40	2.680E - 1	5.721E + 0	2.290E + 1	1.774E + 1	8.403E + 0	7.060E + 0	1.639E + 1	1.344E + 1
45	2.435E - 1	4.153E + 0	1.688E + 1	1.304E + 1	6.111E + 0	5.167E + 0	1.213E + 1	9.950E + 0
50	2.267E - 1	3.067E + 0	1.286E + 1	9.911E + 0	4.609E + 0	3.922E + 0	9.232E + 0	7.580E + 0
55	2.144E - 1	1.225E + 1	1.001E + 1	7.719E + 0	3.573E + 0	3.054E + 0	7.204E + 0	5.920E + 0
60	2.055E - 1	9.833E + 0	7.961E + 0	6.171E + 0	2.845E + 0	2.444E + 0	5.798E + 0	4.770E + 0
65	1.986E - 1	8.050E + 0	6.464E + 0	5.014E + 0	2.299E + 0	1.983E + 0	4.729E + 0	3.895E + 0
70	1.927E - 1	6.590E + 0	5.247E + 0	4.071E + 0	6.415E + 0	7.059E + 0	3.875E + 0	3.196E + 0
75	1.876E - 1	5.553E + 0	4.372E + 0	3.392E + 0	5.401E + 0	5.979E + 0	3.267E + 0	2.699E + 0
80	1.831E - 1	4.686E + 0	3.683E + 0	2.860E + 0	4.574E + 0	5.072E + 0	2.751E + 0	2.276E + 0
85	1.794E - 1	4.056E + 0	3.185E + 0	2.473E + 0	3.973E + 0	4.413E + 0	2.386E + 0	1.978E + 0
90	1.761E - 1	3.437E + 0	2.694E + 0	2.102E + 0	3.395E + 0	3.768E + 0	2.035E + 0	1.691E + 0
95	1.733E - 1	3.023E + 0	2.367E + 0	1.852E + 0	3.000E + 0	3.332E + 0	1.795E + 0	1.494E + 0
100	1.706E - 1	2.613E + 0	2.044E + 0	1.604E + 0	2.607E + 0	2.895E + 0	1.556E + 0	1.299E + 0
110	1.654E - 1	2.043E + 0	1.592E + 0	1.255E + 0	2.056E + 0	2.284E + 0	1.222E + 0	1.025E + 0
120	1.611E - 1	1.647E + 0	1.278E + 0	1.013E + 0	1.675E + 0	1.871E + 0	3.808E + 0	3.130E + 0
130	1.572E - 1	1.356E + 0	1.052E + 0	8.378E - 1	1.370E + 0	1.523E + 0	3.123E + 0	2.572E + 0
140	1.536E - 1	1.116E + 0	8.651E - 1	6.931E - 1	1.129E + 0	1.264E + 0	2.579E + 0	2.127E + 0
150	1.504E - 1	9.443E - 1	7.326E - 1	5.905E - 1	9.569E - 1	1.071E + 0	2.186E + 0	1.807E + 0
160	1.473E - 1	8.088E - 1	6.282E - 1	5.097E - 1	8.206E - 1	9.177E - 1	1.867E + 0	1.546E + 0
170	1.445E - 1	7.109E - 1	5.536E - 1	4.515E - 1	7.221E - 1	8.067E - 1	1.637E + 0	1.359E + 0
180	1.417E - 1	6.147E - 1	4.799E - 1	3.947E - 1	6.258E - 1	6.981E - 1	1.410E + 0	1.173E + 0
190	1.391E - 1	5.501E - 1	4.309E - 1	3.565E - 1	5.606E - 1	6.245E - 1	1.255E + 0	1.046E + 0
200	1.366E - 1	4.858E - 1	3.820E - 1	3.186E - 1	4.956E - 1	5.512E - 1	1.101E + 0	9.201E - 1

^aPolymethylmetacrylate, C₅H₈O₂.

^bComposition data from Hammerstein GR, Miller DW, White DR, et al. Absorbed radiation dose in mammography. *Radiology* 1979;130:485-491.

C.3 MASS ENERGY ATTENUATION COEFFICIENTS FOR SELECTED DETECTOR COMPOUNDS

TABLE C-3. MASS ENERGY ATTENUATION COEFFICIENTS IN CM²/G (DENSITY (ρ) IN G/CM³)

Energy (keV)	Si (elemental) $\rho = 2.33$	Se (elemental) $\rho = 4.79$	BaFBr $\rho = 4.56$	CsI $\rho = 4.51$	Gd ₂ O ₂ S $\rho = 7.34$	YTaO ₄ $\rho = 7.57$	CaWO ₄ $\rho = 6.12$	AgBr $\rho = 6.47$
2	2.651E + 3	3.051E + 3	2.465E + 3	2.106E + 3	2.871E + 3	2.357E + 3	2.750E + 3	2.217E + 3
3	9.542E + 2	1.100E + 3	9.131E + 2	7.921E + 2	1.206E + 3	1.472E + 3	1.305E + 3	8.068E + 2
4	4.410E + 2	5.158E + 2	4.326E + 2	3.766E + 2	5.830E + 2	7.192E + 2	6.408E + 2	9.569E + 2
5	2.390E + 2	2.841E + 2	2.863E + 2	4.955E + 2	3.300E + 2	4.078E + 2	4.323E + 2	5.444E + 2
6	1.440E + 2	1.748E + 2	4.627E + 2	6.024E + 2	2.091E + 2	2.598E + 2	2.765E + 2	3.415E + 2
7	9.275E + 1	1.132E + 2	3.177E + 2	4.073E + 2	1.373E + 2	1.702E + 2	1.827E + 2	2.259E + 2
8	6.278E + 1	7.857E + 1	2.226E + 2	2.874E + 2	3.043E + 2	1.204E + 2	1.298E + 2	1.580E + 2
9	4.450E + 1	5.629E + 1	1.631E + 2	2.125E + 2	2.628E + 2	8.781E + 1	9.495E + 1	1.151E + 2
10	3.270E + 1	4.187E + 1	1.243E + 2	1.618E + 2	2.034E + 2	1.273E + 2	7.182E + 1	8.664E + 1
12	1.940E + 1	2.549E + 1	7.700E + 1	1.000E + 2	1.293E + 2	1.163E + 2	1.187E + 2	5.284E + 1
14	1.198E + 1	7.037E + 1	6.520E + 1	6.582E + 1	8.612E + 1	7.936E + 1	9.304E + 1	5.622E + 1
16	7.983E + 0	5.401E + 1	4.710E + 1	4.586E + 1	6.061E + 1	5.783E + 1	6.709E + 1	4.147E + 1
18	5.579E + 0	4.222E + 1	3.512E + 1	3.319E + 1	4.437E + 1	5.073E + 1	4.978E + 1	3.147E + 1
20	4.047E + 0	3.316E + 1	2.694E + 1	2.487E + 1	3.362E + 1	3.954E + 1	3.806E + 1	2.446E + 1
25	2.049E + 0	1.942E + 1	1.510E + 1	1.339E + 1	1.838E + 1	2.297E + 1	2.135E + 1	1.391E + 1
30	1.161E + 0	1.238E + 1	9.364E + 0	8.119E + 0	1.126E + 1	1.465E + 1	1.334E + 1	1.511E + 1
35	7.182E - 1	8.225E + 0	6.611E + 0	8.372E + 0	7.344E + 0	9.783E + 0	8.771E + 0	1.114E + 1
40	4.762E - 1	5.753E + 0	7.447E + 0	9.367E + 0	5.060E + 0	6.906E + 0	6.119E + 0	8.323E + 0
45	3.332E - 1	4.187E + 0	6.192E + 0	7.877E + 0	3.652E + 0	5.036E + 0	4.444E + 0	6.355E + 0
50	2.436E - 1	3.146E + 0	5.131E + 0	6.580E + 0	2.735E + 0	3.798E + 0	3.344E + 0	4.958E + 0
55	1.838E - 1	2.403E + 0	4.258E + 0	5.484E + 0	4.326E + 0	2.939E + 0	2.580E + 0	3.927E + 0
60	1.444E - 1	1.888E + 0	3.572E + 0	4.598E + 0	3.930E + 0	2.334E + 0	2.048E + 0	3.188E + 0
65	1.168E - 1	1.515E + 0	3.014E + 0	3.888E + 0	3.528E + 0	1.890E + 0	1.653E + 0	2.604E + 0
70	9.516E - 2	1.213E + 0	2.523E + 0	3.263E + 0	3.108E + 0	2.258E + 0	2.062E + 0	2.125E + 0
75	8.123E - 2	9.988E - 1	2.153E + 0	2.789E + 0	2.773E + 0	2.090E + 0	1.980E + 0	1.775E + 0
80	6.910E - 2	8.318E - 1	1.849E + 0	2.400E + 0	2.452E + 0	1.910E + 0	1.858E + 0	1.498E + 0
85	6.169E - 2	7.029E - 1	1.624E + 0	2.112E + 0	2.207E + 0	1.763E + 0	1.753E + 0	1.294E + 0
90	5.427E - 2	5.974E - 1	1.394E + 0	1.808E + 0	1.929E + 0	1.584E + 0	1.596E + 0	1.097E + 0
95	4.976E - 2	5.214E - 1	1.240E + 0	1.607E + 0	1.744E + 0	1.461E + 0	1.490E + 0	9.646E - 1
100	4.524E - 2	4.454E - 1	1.079E + 0	1.397E + 0	1.539E + 0	1.313E + 0	1.353E + 0	8.299E - 1
110	3.980E - 2	3.417E - 1	8.492E - 1	1.100E + 0	1.247E + 0	1.095E + 0	1.145E + 0	6.426E - 1
120	3.638E - 2	2.730E - 1	6.853E - 1	8.860E - 1	1.029E + 0	9.292E - 1	9.872E - 1	5.129E - 1
130	3.392E - 2	2.205E - 1	5.645E - 1	7.300E - 1	8.624E - 1	7.831E - 1	8.354E - 1	4.178E - 1
140	3.212E - 2	1.790E - 1	4.618E - 1	5.961E - 1	7.167E - 1	6.588E - 1	7.138E - 1	3.389E - 1
150	3.111E - 2	1.501E - 1	3.884E - 1	5.006E - 1	6.097E - 1	5.667E - 1	6.176E - 1	2.834E - 1
160	3.033E - 2	1.281E - 1	3.304E - 1	4.249E - 1	5.237E - 1	4.910E - 1	5.376E - 1	2.402E - 1
170	2.976E - 2	1.120E - 1	2.890E - 1	3.712E - 1	4.612E - 1	4.359E - 1	4.788E - 1	2.093E - 1
180	2.959E - 2	9.754E - 2	2.473E - 1	3.163E - 1	3.966E - 1	3.779E - 1	4.164E - 1	1.792E - 1
190	2.951E - 2	8.772E - 2	2.199E - 1	2.802E - 1	3.539E - 1	3.395E - 1	3.749E - 1	1.593E - 1
200	2.944E - 2	7.790E - 2	1.920E - 1	2.440E - 1	3.096E - 1	2.990E - 1	3.308E - 1	1.393E - 1

C.4 MAMMOGRAPHY SPECTRA: Mo/Mo**TABLE C-4. SPECTRA FOR MOLYBDENUM ANODE WITH 30 μ M MOLYBDENUM FILTER AND 3 MM LUCITE FILTER BY HALF VALUE LAYER IN MM Al^a**

Energy (keV)	24 kVp 0.309 mm	25 kVp 0.322 mm	26 kVp 0.334 mm	27 kVp 0.345 mm	28 kVp 0.356 mm	30 kVp 0.375 mm	32 kVp 0.391 mm	34 kVp 0.406 mm
5.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
5.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
6.0	6.744E - 1	6.009E - 1	5.398E - 1	4.882E - 1	4.442E - 1	3.732E - 1	3.193E - 1	2.776E - 1
6.5	1.853E + 1	1.662E + 1	1.502E + 1	1.367E + 1	1.252E + 1	1.064E + 1	9.208E + 0	8.088E + 0
7.0	2.537E + 2	2.277E + 2	2.059E + 2	1.874E + 2	1.715E + 2	1.456E + 2	1.256E + 2	1.098E + 2
7.5	1.506E + 3	1.351E + 3	1.220E + 3	1.109E + 3	1.013E + 3	8.549E + 2	7.323E + 2	6.348E + 2
8.0	7.382E + 3	6.683E + 3	6.095E + 3	5.592E + 3	5.158E + 3	4.445E + 3	3.890E + 3	3.448E + 3
8.5	2.215E + 4	2.012E + 4	1.841E + 4	1.695E + 4	1.568E + 4	1.360E + 4	1.196E + 4	1.065E + 4
9.0	5.608E + 4	5.097E + 4	4.666E + 4	4.297E + 4	3.978E + 4	3.451E + 4	3.041E + 4	2.714E + 4
9.5	1.069E + 5	9.751E + 4	8.950E + 4	8.263E + 4	7.665E + 4	6.675E + 4	5.896E + 4	5.270E + 4
10.0	2.013E + 5	1.841E + 5	1.695E + 5	1.570E + 5	1.461E + 5	1.280E + 5	1.138E + 5	1.023E + 5
10.5	3.004E + 5	2.753E + 5	2.540E + 5	2.357E + 5	2.197E + 5	1.931E + 5	1.722E + 5	1.552E + 5
11.0	4.555E + 5	4.182E + 5	3.863E + 5	3.587E + 5	3.347E + 5	2.948E + 5	2.635E + 5	2.385E + 5
11.5	5.970E + 5	5.501E + 5	5.100E + 5	4.753E + 5	4.450E + 5	3.944E + 5	3.543E + 5	3.219E + 5
12.0	7.587E + 5	7.011E + 5	6.516E + 5	6.086E + 5	5.708E + 5	5.075E + 5	4.571E + 5	4.164E + 5
12.5	9.430E + 5	8.752E + 5	8.167E + 5	7.655E + 5	7.203E + 5	6.441E + 5	5.828E + 5	5.327E + 5
13.0	1.091E + 6	1.015E + 6	9.482E + 5	8.895E + 5	8.374E + 5	7.493E + 5	6.786E + 5	6.212E + 5
13.5	1.232E + 6	1.151E + 6	1.078E + 6	1.015E + 6	9.580E + 5	8.612E + 5	7.832E + 5	7.196E + 5
14.0	1.383E + 6	1.297E + 6	1.220E + 6	1.152E + 6	1.090E + 6	9.846E + 5	8.988E + 5	8.284E + 5
14.5	1.475E + 6	1.390E + 6	1.313E + 6	1.244E + 6	1.181E + 6	1.072E + 6	9.819E + 5	9.072E + 5
15.0	1.563E + 6	1.481E + 6	1.406E + 6	1.338E + 6	1.275E + 6	1.165E + 6	1.072E + 6	9.945E + 5
15.5	1.627E + 6	1.552E + 6	1.481E + 6	1.415E + 6	1.354E + 6	1.245E + 6	1.153E + 6	1.074E + 6
16.0	1.684E + 6	1.618E + 6	1.553E + 6	1.491E + 6	1.432E + 6	1.326E + 6	1.233E + 6	1.154E + 6
16.5	1.692E + 6	1.639E + 6	1.585E + 6	1.530E + 6	1.478E + 6	1.379E + 6	1.291E + 6	1.216E + 6
17.0	2.100E + 6	2.173E + 6	2.235E + 6	2.287E + 6	2.331E + 6	2.402E + 6	2.456E + 6	2.496E + 6
17.5	5.682E + 6	6.726E + 6	7.696E + 6	8.593E + 6	9.415E + 6	1.085E + 7	1.204E + 7	1.302E + 7
18.0	1.713E + 6	1.730E + 6	1.733E + 6	1.728E + 6	1.716E + 6	1.682E + 6	1.643E + 6	1.604E + 6
18.5	1.516E + 6	1.528E + 6	1.520E + 6	1.501E + 6	1.474E + 6	1.409E + 6	1.341E + 6	1.277E + 6
19.0	1.463E + 6	1.499E + 6	1.509E + 6	1.502E + 6	1.484E + 6	1.431E + 6	1.371E + 6	1.313E + 6
19.5	2.211E + 6	2.490E + 6	2.722E + 6	2.917E + 6	3.085E + 6	3.356E + 6	3.567E + 6	3.736E + 6
20.0	1.902E + 5	2.100E + 5	2.250E + 5	2.367E + 5	2.459E + 5	2.589E + 5	2.672E + 5	2.725E + 5
20.5	1.192E + 5	1.302E + 5	1.364E + 5	1.394E + 5	1.401E + 5	1.376E + 5	1.323E + 5	1.261E + 5
21.0	1.135E + 5	1.287E + 5	1.377E + 5	1.424E + 5	1.442E + 5	1.426E + 5	1.376E + 5	1.314E + 5
21.5	1.073E + 5	1.275E + 5	1.403E + 5	1.479E + 5	1.520E + 5	1.536E + 5	1.505E + 5	1.452E + 5
22.0	1.019E + 5	1.268E + 5	1.431E + 5	1.534E + 5	1.595E + 5	1.639E + 5	1.624E + 5	1.581E + 5
22.5	7.880E + 4	1.136E + 5	1.369E + 5	1.523E + 5	1.621E + 5	1.710E + 5	1.719E + 5	1.688E + 5
23.0	5.348E + 4	9.647E + 4	1.263E + 5	1.468E + 5	1.608E + 5	1.758E + 5	1.805E + 5	1.796E + 5
23.5	2.190E + 4	7.741E + 4	1.157E + 5	1.420E + 5	1.598E + 5	1.792E + 5	1.860E + 5	1.865E + 5
24.0	8.645E + 3	6.865E + 4	1.109E + 5	1.407E + 5	1.617E + 5	1.861E + 5	1.966E + 5	1.996E + 5
24.5	0.000E + 0	4.441E + 4	9.177E + 4	1.268E + 5	1.529E + 5	1.863E + 5	2.038E + 5	2.115E + 5
25.0	0.000E + 0	0.000E + 0	6.238E + 4	1.090E + 5	1.438E + 5	1.889E + 5	2.131E + 5	2.245E + 5
25.5	0.000E + 0	0.000E + 0	3.626E + 4	8.848E + 4	1.279E + 5	1.801E + 5	2.093E + 5	2.244E + 5
26.0	0.000E + 0	0.000E + 0	1.423E + 4	7.220E + 4	1.162E + 5	1.753E + 5	2.092E + 5	2.276E + 5
26.5	0.000E + 0	0.000E + 0	0.000E + 0	3.935E + 4	9.269E + 4	1.649E + 5	2.071E + 5	2.307E + 5
27.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.173E + 4	1.506E + 5	2.020E + 5	2.277E + 5
27.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.923E + 4	1.356E + 5	1.940E + 5	2.288E + 5
28.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.339E + 4	1.194E + 5	1.846E + 5	2.245E + 5
28.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	9.360E + 4	1.701E + 5	2.175E + 5
29.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	6.517E + 4	1.503E + 5	2.039E + 5

(continued)

TABLE C-4 (continued).

Energy (keV)	24 kVp 0.309 mm	25 kVp 0.322 mm	26 kVp 0.334 mm	27 kVp 0.345 mm	28 kVp 0.356 mm	30 kVp 0.375 mm	32 kVp 0.391 mm	34 kVp 0.406 mm
29.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.788E + 4	1.341E + 5	1.949E + 5
30.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.506E + 4	1.173E + 5	1.830E + 5
30.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.646E + 4	1.665E + 5
31.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.501E + 4	1.487E + 5
31.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.225E + 4	1.283E + 5
32.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.103E + 4	1.132E + 5
32.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.760E + 4
33.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.814E + 4
33.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.120E + 4
34.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.296E + 4
34.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
35.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0

*Each spectrum is normalized to 1 R, and individual table entries are in the units of photons/mm² per Energy interval.

C.5 MAMMOGRAPHY SPECTRA: Mo/Rh**TABLE C-5. SPECTRA FOR MOLYBDENUM ANODE WITH 25 μ M RHODIUM FILTER AND 3 MM LUCITE FILTER BY HALF VALUE LAYER IN MM Al^a**

Energy (keV)	26 kVp 0.414 mm	27 kVp 0.426 mm	28 kVp 0.437 mm	29 kVp 0.446 mm	30 kVp 0.454 mm	31 kVp 0.461 mm	32 kVp 0.468 mm	34 kVp 0.479 mm
5.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
5.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
6.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
6.5	3.583E - 1	3.209E - 1	2.901E - 1	2.644E - 1	2.426E - 1	2.239E - 1	2.077E - 1	1.812E - 1
7.0	1.071E + 1	9.592E + 0	8.671E + 0	7.897E + 0	7.239E + 0	6.672E + 0	6.180E + 0	5.367E + 0
7.5	1.085E + 2	9.698E + 1	8.748E + 1	7.947E + 1	7.263E + 1	6.673E + 1	6.157E + 1	5.303E + 1
8.0	8.389E + 2	7.573E + 2	6.899E + 2	6.331E + 2	5.846E + 2	5.428E + 2	5.064E + 2	4.459E + 2
8.5	3.268E + 3	2.960E + 3	2.705E + 3	2.490E + 3	2.307E + 3	2.148E + 3	2.009E + 3	1.777E + 3
9.0	1.160E + 4	1.051E + 4	9.610E + 3	8.850E + 3	8.200E + 3	7.640E + 3	7.150E + 3	6.339E + 3
9.5	2.754E + 4	2.502E + 4	2.292E + 4	2.115E + 4	1.963E + 4	1.831E + 4	1.716E + 4	1.523E + 4
10.0	6.448E + 4	5.874E + 4	5.399E + 4	4.997E + 4	4.652E + 4	4.354E + 4	4.093E + 4	3.657E + 4
10.5	1.120E + 5	1.022E + 5	9.412E + 4	8.726E + 4	8.137E + 4	7.626E + 4	7.178E + 4	6.429E + 4
11.0	1.971E + 5	1.801E + 5	1.659E + 5	1.540E + 5	1.437E + 5	1.349E + 5	1.272E + 5	1.143E + 5
11.5	2.896E + 5	2.656E + 5	2.456E + 5	2.286E + 5	2.141E + 5	2.014E + 5	1.903E + 5	1.718E + 5
12.0	4.047E + 5	3.719E + 5	3.445E + 5	3.212E + 5	3.012E + 5	2.838E + 5	2.685E + 5	2.430E + 5
12.5	5.527E + 5	5.097E + 5	4.737E + 5	4.431E + 5	4.165E + 5	3.934E + 5	3.730E + 5	3.387E + 5
13.0	6.827E + 5	6.301E + 5	5.859E + 5	5.482E + 5	5.155E + 5	4.871E + 5	4.621E + 5	4.202E + 5
13.5	8.277E + 5	7.662E + 5	7.144E + 5	6.700E + 5	6.316E + 5	5.981E + 5	5.684E + 5	5.188E + 5
14.0	9.923E + 5	9.214E + 5	8.615E + 5	8.100E + 5	7.652E + 5	7.260E + 5	6.913E + 5	6.329E + 5
14.5	1.121E + 6	1.045E + 6	9.801E + 5	9.239E + 5	8.748E + 5	8.316E + 5	7.931E + 5	7.279E + 5
15.0	1.255E + 6	1.174E + 6	1.106E + 6	1.046E + 6	9.930E + 5	9.465E + 5	9.048E + 5	8.337E + 5
15.5	1.371E + 6	1.289E + 6	1.219E + 6	1.157E + 6	1.102E + 6	1.053E + 6	1.010E + 6	9.345E + 5
16.0	1.491E + 6	1.409E + 6	1.337E + 6	1.273E + 6	1.217E + 6	1.166E + 6	1.120E + 6	1.041E + 6
16.5	1.565E + 6	1.487E + 6	1.418E + 6	1.357E + 6	1.301E + 6	1.251E + 6	1.206E + 6	1.128E + 6
17.0	2.269E + 6	2.285E + 6	2.300E + 6	2.316E + 6	2.331E + 6	2.345E + 6	2.358E + 6	2.382E + 6
17.5	8.002E + 6	8.789E + 6	9.512E + 6	1.017E + 7	1.078E + 7	1.134E + 7	1.184E + 7	1.272E + 7
18.0	1.847E + 6	1.811E + 6	1.777E + 6	1.744E + 6	1.713E + 6	1.684E + 6	1.656E + 6	1.606E + 6
18.5	1.640E + 6	1.593E + 6	1.545E + 6	1.498E + 6	1.453E + 6	1.410E + 6	1.369E + 6	1.294E + 6
19.0	1.652E + 6	1.617E + 6	1.578E + 6	1.538E + 6	1.497E + 6	1.457E + 6	1.420E + 6	1.350E + 6
19.5	3.021E + 6	3.186E + 6	3.327E + 6	3.451E + 6	3.559E + 6	3.657E + 6	3.745E + 6	3.896E + 6
20.0	1.939E + 6	2.007E + 6	2.060E + 6	2.100E + 6	2.132E + 6	2.158E + 6	2.178E + 6	2.206E + 6
20.5	1.163E + 6	1.169E + 6	1.161E + 6	1.144E + 6	1.121E + 6	1.095E + 6	1.067E + 6	1.010E + 6
21.0	1.116E + 6	1.135E + 6	1.136E + 6	1.124E + 6	1.105E + 6	1.081E + 6	1.055E + 6	1.001E + 6
21.5	1.046E + 6	1.085E + 6	1.102E + 6	1.103E + 6	1.095E + 6	1.080E + 6	1.061E + 6	1.017E + 6
22.0	9.760E + 5	1.029E + 6	1.057E + 6	1.068E + 6	1.068E + 6	1.060E + 6	1.047E + 6	1.013E + 6
22.5	8.696E + 5	9.517E + 5	1.001E + 6	1.027E + 6	1.038E + 6	1.039E + 6	1.033E + 6	1.007E + 6
23.0	7.481E + 5	8.560E + 5	9.261E + 5	9.700E + 5	9.955E + 5	1.008E + 6	1.011E + 6	9.997E + 5
23.5	9.160E + 4	1.106E + 5	1.230E + 5	1.309E + 5	1.356E + 5	1.383E + 5	1.394E + 5	1.388E + 5
24.0	8.910E + 4	1.112E + 5	1.262E + 5	1.363E + 5	1.429E + 5	1.470E + 5	1.494E + 5	1.507E + 5
24.5	7.499E + 4	1.020E + 5	1.214E + 5	1.354E + 5	1.455E + 5	1.526E + 5	1.575E + 5	1.624E + 5
25.0	5.192E + 4	8.925E + 4	1.163E + 5	1.360E + 5	1.503E + 5	1.606E + 5	1.677E + 5	1.755E + 5
25.5	3.058E + 4	7.342E + 4	1.048E + 5	1.280E + 5	1.451E + 5	1.578E + 5	1.669E + 5	1.778E + 5
26.0	1.217E + 4	6.076E + 4	9.661E + 4	1.233E + 5	1.433E + 5	1.582E + 5	1.692E + 5	1.829E + 5
26.5	0.000E + 0	3.362E + 4	7.823E + 4	1.117E + 5	1.368E + 5	1.558E + 5	1.701E + 5	1.883E + 5
27.0	0.000E + 0	0.000E + 0	4.428E + 4	9.249E + 4	1.268E + 5	1.512E + 5	1.683E + 5	1.884E + 5
27.5	0.000E + 0	0.000E + 0	3.407E + 4	8.039E + 4	1.158E + 5	1.431E + 5	1.640E + 5	1.921E + 5
28.0	0.000E + 0	0.000E + 0	1.180E + 4	6.355E + 4	1.034E + 5	1.343E + 5	1.583E + 5	1.912E + 5
28.5	0.000E + 0	0.000E + 0	0.000E + 0	3.479E + 4	8.203E + 4	1.188E + 5	1.475E + 5	1.874E + 5
29.0	0.000E + 0	0.000E + 0	0.000E + 0	4.751E + 3	5.775E + 4	9.925E + 4	1.319E + 5	1.777E + 5

(continued)

TABLE C-5 (continued).

Energy (keV)	26 kVp 0.414 mm	27 kVp 0.426 mm	28 kVp 0.437 mm	29 kVp 0.446 mm	30 kVp 0.454 mm	31 kVp 0.461 mm	32 kVp 0.468 mm	34 kVp 0.479 mm
29.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.396E + 4	8.148E + 4	1.189E + 5	1.718E + 5
30.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.366E + 4	6.473E + 4	1.052E + 5	1.631E + 5
30.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	2.781E + 4	7.847E + 4	1.502E + 5
31.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.057E + 4	1.358E + 5
31.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.002E + 4	1.186E + 5
32.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.040E + 4	1.060E + 5
32.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.302E + 4
33.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.576E + 4
33.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.026E + 4
34.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.271E + 4
34.5	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
35.0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0

^aEach spectrum is normalized to 1 R, and individual table entries are in the units of photons/mm² per energy interval.

C.6 MAMMOGRAPHY SPECTRA: Rh/Rh

TABLE C-6. SPECTRA FOR RHODIUM ANODE WITH 25 μM RHODIUM FILTER AND 3 MM LUCITE FILTER BY HALF VALUE LAYER IN MM Al^a

Energy (keV)	26 kVp 0.375 mm	28 kVp 0.409 mm	29 kVp 0.424 mm	30 kVp 0.439 mm	31 kVp 0.453 mm	32 kVp 0.466 mm	33 kVp 0.479 mm	34 kVp 0.490 mm
5.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
5.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
6.2	3.850E - 1	3.179E - 1	2.914E - 1	2.685E - 1	2.484E - 1	2.307E - 1	2.150E - 1	2.011E - 1
6.7	9.756E + 0	8.060E + 0	7.387E + 0	6.804E + 0	6.295E + 0	5.846E + 0	5.449E + 0	5.098E + 0
7.2	1.307E + 2	1.087E + 2	9.998E + 1	9.239E + 1	8.574E + 1	7.987E + 1	7.466E + 1	7.004E + 1
7.7	8.991E + 2	7.532E + 2	6.952E + 2	6.447E + 2	6.004E + 2	5.612E + 2	5.265E + 2	4.955E + 2
8.2	4.191E + 3	3.527E + 3	3.262E + 3	3.030E + 3	2.827E + 3	2.647E + 3	2.487E + 3	2.344E + 3
8.7	1.365E + 4	1.154E + 4	1.070E + 4	9.964E + 3	9.316E + 3	8.741E + 3	8.228E + 3	7.770E + 3
9.2	3.598E + 4	3.061E + 4	2.845E + 4	2.657E + 4	2.490E + 4	2.343E + 4	2.211E + 4	2.093E + 4
9.7	7.610E + 4	6.516E + 4	6.074E + 4	5.686E + 4	5.342E + 4	5.034E + 4	4.758E + 4	4.510E + 4
10.2	1.459E + 5	1.250E + 5	1.166E + 5	1.092E + 5	1.026E + 5	9.669E + 4	9.143E + 4	8.671E + 4
10.7	2.425E + 5	2.097E + 5	1.963E + 5	1.846E + 5	1.742E + 5	1.649E + 5	1.565E + 5	1.491E + 5
11.2	3.679E + 5	3.188E + 5	2.989E + 5	2.813E + 5	2.658E + 5	2.519E + 5	2.394E + 5	2.283E + 5
11.7	5.060E + 5	4.403E + 5	4.134E + 5	3.898E + 5	3.688E + 5	3.500E + 5	3.332E + 5	3.181E + 5
12.2	6.557E + 5	5.740E + 5	5.405E + 5	5.109E + 5	4.845E + 5	4.608E + 5	4.395E + 5	4.204E + 5
12.7	8.247E + 5	7.262E + 5	6.856E + 5	6.496E + 5	6.173E + 5	5.882E + 5	5.619E + 5	5.382E + 5
13.2	9.870E + 5	8.724E + 5	8.247E + 5	7.822E + 5	7.441E + 5	7.097E + 5	6.787E + 5	6.508E + 5
13.7	1.142E + 6	1.016E + 6	9.628E + 5	9.155E + 5	8.729E + 5	8.342E + 5	7.990E + 5	7.671E + 5
14.2	1.296E + 6	1.159E + 6	1.101E + 6	1.049E + 6	1.002E + 6	9.595E + 5	9.207E + 5	8.854E + 5
14.7	1.414E + 6	1.273E + 6	1.212E + 6	1.157E + 6	1.107E + 6	1.061E + 6	1.019E + 6	9.810E + 5
15.2	1.532E + 6	1.383E + 6	1.319E + 6	1.261E + 6	1.207E + 6	1.159E + 6	1.114E + 6	1.074E + 6
15.7	1.637E + 6	1.501E + 6	1.440E + 6	1.383E + 6	1.331E + 6	1.281E + 6	1.235E + 6	1.192E + 6
16.2	1.709E + 6	1.566E + 6	1.502E + 6	1.443E + 6	1.388E + 6	1.337E + 6	1.291E + 6	1.248E + 6
16.7	1.753E + 6	1.615E + 6	1.551E + 6	1.492E + 6	1.438E + 6	1.387E + 6	1.341E + 6	1.300E + 6
17.2	1.867E + 6	1.768E + 6	1.719E + 6	1.674E + 6	1.630E + 6	1.590E + 6	1.552E + 6	1.518E + 6
17.7	1.886E + 6	1.817E + 6	1.780E + 6	1.742E + 6	1.706E + 6	1.671E + 6	1.638E + 6	1.607E + 6
18.2	1.784E + 6	1.698E + 6	1.651E + 6	1.605E + 6	1.560E + 6	1.517E + 6	1.475E + 6	1.437E + 6
18.7	1.702E + 6	1.653E + 6	1.621E + 6	1.587E + 6	1.552E + 6	1.518E + 6	1.484E + 6	1.451E + 6
19.2	1.711E + 6	1.676E + 6	1.645E + 6	1.612E + 6	1.576E + 6	1.541E + 6	1.506E + 6	1.473E + 6
19.7	1.920E + 6	2.093E + 6	2.158E + 6	2.213E + 6	2.261E + 6	2.303E + 6	2.340E + 6	2.375E + 6
20.2	3.744E + 6	5.797E + 6	6.781E + 6	7.716E + 6	8.597E + 6	9.418E + 6	1.018E + 7	1.089E + 7
20.7	1.561E + 6	1.740E + 6	1.803E + 6	1.854E + 6	1.894E + 6	1.926E + 6	1.950E + 6	1.969E + 6
21.2	1.360E + 6	1.448E + 6	1.460E + 6	1.460E + 6	1.452E + 6	1.438E + 6	1.421E + 6	1.402E + 6
21.7	1.255E + 6	1.382E + 6	1.407E + 6	1.416E + 6	1.415E + 6	1.407E + 6	1.394E + 6	1.379E + 6
22.2	1.142E + 6	1.320E + 6	1.368E + 6	1.399E + 6	1.418E + 6	1.427E + 6	1.429E + 6	1.426E + 6
22.7	1.447E + 6	2.114E + 6	2.392E + 6	2.640E + 6	2.862E + 6	3.061E + 6	3.239E + 6	3.401E + 6
23.2	1.799E + 5	2.449E + 5	2.670E + 5	2.846E + 5	2.985E + 5	3.097E + 5	3.188E + 5	3.264E + 5
23.7	1.294E + 5	1.770E + 5	1.909E + 5	2.004E + 5	2.067E + 5	2.106E + 5	2.126E + 5	2.132E + 5
24.2	1.268E + 5	1.780E + 5	1.932E + 5	2.039E + 5	2.112E + 5	2.158E + 5	2.186E + 5	2.199E + 5
24.7	1.025E + 5	1.720E + 5	1.929E + 5	2.078E + 5	2.181E + 5	2.249E + 5	2.291E + 5	2.313E + 5
25.2	7.466E + 4	1.607E + 5	1.867E + 5	2.054E + 5	2.186E + 5	2.277E + 5	2.338E + 5	2.375E + 5
25.7	3.353E + 4	1.372E + 5	1.700E + 5	1.942E + 5	2.120E + 5	2.248E + 5	2.338E + 5	2.398E + 5
26.2	0.000E + 0	1.307E + 5	1.641E + 5	1.893E + 5	2.084E + 5	2.228E + 5	2.335E + 5	2.413E + 5
26.7	0.000E + 0	1.091E + 5	1.479E + 5	1.777E + 5	2.007E + 5	2.183E + 5	2.319E + 5	2.422E + 5
27.2	0.000E + 0	7.349E + 4	1.215E + 5	1.583E + 5	1.866E + 5	2.083E + 5	2.248E + 5	2.374E + 5
27.7	0.000E + 0	2.888E + 4	9.244E + 4	1.412E + 5	1.785E + 5	2.071E + 5	2.289E + 5	2.453E + 5
28.2	0.000E + 0	0.000E + 0	8.383E + 4	1.301E + 5	1.661E + 5	1.942E + 5	2.161E + 5	2.332E + 5
28.7	0.000E + 0	0.000E + 0	3.655E + 4	9.455E + 4	1.400E + 5	1.758E + 5	2.039E + 5	2.260E + 5
29.2	0.000E + 0	0.000E + 0	0.000E + 0	6.893E + 4	1.180E + 5	1.570E + 5	1.881E + 5	2.130E + 5
29.7	0.000E + 0	0.000E + 0	0.000E + 0	3.036E + 4	9.084E + 4	1.383E + 5	1.755E + 5	2.046E + 5
30.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	6.420E + 4	1.167E + 5	1.580E + 5	1.905E + 5

(continued)

TABLE C-6 (continued)

Energy (keV)	26 kVp 0.375 mm	28 kVp 0.409 mm	29 kVp 0.424 mm	30 kVp 0.439 mm	31 kVp 0.453 mm	32 kVp 0.466 mm	33 kVp 0.479 mm	34 kVp 0.490 mm
30.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.114E + 4	8.703E + 4	1.318E + 5	1.678E + 5
31.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	6.719E + 4	1.144E + 5	1.526E + 5
31.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.817E + 4	8.530E + 4	1.248E + 5
32.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	6.084E + 4	1.085E + 5
32.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.084E + 4	9.373E + 4
33.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	6.179E + 4
33.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	2.902E + 4
34.2	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
34.7	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0

^aEach spectrum is normalized to 1 R, and individual table entries are in the units of photons/mm² per energy interval.

C.7 GENERAL DIAGNOSTIC SPECTRA: W/AI**TABLE C-7. SPECTRA FOR TUNGSTEN ANODE (5% RIPPLE) WITH 2.5 MM ADDED AI FILTRATION BY HALF VALUE LAYER IN MM Al^a**

Energy (keV)	50 kVp ^b 2.54 mm	60 kVp ^b 3.21 mm	70 kVp 4.18 mm	80 kVp 4.82 mm	90 kVp 5.41 mm	100 kVp 5.94 mm	110 kVp 6.43 mm	120 kVp 6.89 mm
10	9.519E - 1	6.141E - 1	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0
12	5.043E + 2	2.947E + 2	8.738E + 0	6.341E + 0	5.021E + 0	4.290E + 0	3.905E + 0	3.737E + 0
14	3.018E + 4	1.738E + 4	1.591E + 3	1.140E + 3	8.773E + 2	7.131E + 2	6.041E + 2	5.284E + 2
16	3.974E + 5	2.388E + 5	4.377E + 4	3.206E + 4	2.494E + 4	2.027E + 4	1.698E + 4	1.453E + 4
18	1.963E + 6	1.224E + 6	3.439E + 5	2.561E + 5	2.013E + 5	1.645E + 5	1.380E + 5	1.176E + 5
20	7.475E + 5	4.709E + 5	1.776E + 5	1.324E + 5	1.040E + 5	8.491E + 4	7.115E + 4	6.067E + 4
22	2.015E + 6	1.308E + 6	6.123E + 5	4.624E + 5	3.665E + 5	3.010E + 5	2.533E + 5	2.165E + 5
24	4.358E + 6	2.946E + 6	1.622E + 6	1.243E + 6	9.962E + 5	8.258E + 5	7.013E + 5	6.057E + 5
26	7.345E + 6	5.209E + 6	3.234E + 6	2.515E + 6	2.034E + 6	1.697E + 6	1.449E + 6	1.258E + 6
28	1.078E + 7	8.005E + 6	5.487E + 6	4.327E + 6	3.533E + 6	2.969E + 6	2.550E + 6	2.227E + 6
30	1.369E + 7	1.047E + 7	7.748E + 6	6.217E + 6	5.151E + 6	4.386E + 6	3.809E + 6	3.357E + 6
32	1.575E + 7	1.250E + 7	9.861E + 6	8.025E + 6	6.724E + 6	5.779E + 6	5.063E + 6	4.499E + 6
34	1.667E + 7	1.396E + 7	1.169E + 7	9.677E + 6	8.207E + 6	7.120E + 6	6.286E + 6	5.622E + 6
36	1.658E + 7	1.485E + 7	1.313E + 7	1.103E + 7	9.443E + 6	8.251E + 6	7.330E + 6	6.595E + 6
38	1.556E + 7	1.516E + 7	1.414E + 7	1.207E + 7	1.044E + 7	9.193E + 6	8.210E + 6	7.420E + 6
40	1.443E + 7	1.507E + 7	1.472E + 7	1.278E + 7	1.119E + 7	9.940E + 6	8.937E + 6	8.115E + 6
42	1.231E + 7	1.438E + 7	1.476E + 7	1.307E + 7	1.158E + 7	1.037E + 7	9.378E + 6	8.552E + 6
44	9.331E + 6	1.323E + 7	1.449E + 7	1.316E + 7	1.184E + 7	1.072E + 7	9.773E + 6	8.968E + 6
46	5.850E + 6	1.206E + 7	1.415E + 7	1.312E + 7	1.191E + 7	1.084E + 7	9.926E + 6	9.150E + 6
48	2.039E + 6	1.034E + 7	1.336E + 7	1.281E + 7	1.185E + 7	1.090E + 7	1.005E + 7	9.298E + 6
50	1.758E + 5	9.054E + 6	1.267E + 7	1.241E + 7	1.158E + 7	1.071E + 7	9.923E + 6	9.230E + 6
52	0.000E + 0	7.328E + 6	1.156E + 7	1.173E + 7	1.115E + 7	1.043E + 7	9.735E + 6	9.106E + 6
54	0.000E + 0	5.446E + 6	1.039E + 7	1.105E + 7	1.072E + 7	1.016E + 7	9.557E + 6	8.992E + 6
56	0.000E + 0	3.356E + 6	8.945E + 6	1.054E + 7	1.096E + 7	1.096E + 7	1.077E + 7	1.050E + 7
58	0.000E + 0	1.280E + 6	7.462E + 6	1.107E + 7	1.339E + 7	1.495E + 7	1.602E + 7	1.672E + 7
60	0.000E + 0	1.162E + 5	6.226E + 6	1.059E + 7	1.353E + 7	1.560E + 7	1.707E + 7	1.808E + 7
62	0.000E + 0	0.000E + 0	5.038E + 6	7.801E + 6	8.753E + 6	9.012E + 6	8.982E + 6	8.825E + 6
64	0.000E + 0	0.000E + 0	3.672E + 6	6.724E + 6	7.690E + 6	7.875E + 6	7.750E + 6	7.506E + 6
66	0.000E + 0	0.000E + 0	2.179E + 6	5.965E + 6	7.511E + 6	8.131E + 6	8.323E + 6	8.299E + 6
68	0.000E + 0	0.000E + 0	8.955E + 5	5.467E + 6	7.902E + 6	9.282E + 6	1.009E + 7	1.054E + 7
70	0.000E + 0	0.000E + 0	7.412E + 4	4.031E + 6	5.836E + 6	6.534E + 6	6.789E + 6	6.871E + 6
72	0.000E + 0	0.000E + 0	0.000E + 0	2.782E + 6	4.491E + 6	5.115E + 6	5.254E + 6	5.162E + 6
74	0.000E + 0	0.000E + 0	0.000E + 0	1.946E + 6	3.935E + 6	4.655E + 6	4.832E + 6	4.772E + 6
76	0.000E + 0	0.000E + 0	0.000E + 0	1.219E + 6	3.457E + 6	4.325E + 6	4.600E + 6	4.612E + 6
78	0.000E + 0	0.000E + 0	0.000E + 0	4.668E + 5	2.872E + 6	3.893E + 6	4.265E + 6	4.343E + 6
80	0.000E + 0	0.000E + 0	0.000E + 0	1.384E + 4	2.283E + 6	3.465E + 6	3.935E + 6	4.081E + 6
82	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.784E + 6	3.048E + 6	3.609E + 6	3.831E + 6
84	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.229E + 6	2.668E + 6	3.348E + 6	3.624E + 6
86	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	7.288E + 5	2.267E + 6	3.027E + 6	3.369E + 6
88	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	2.904E + 5	1.935E + 6	2.794E + 6	3.189E + 6
90	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	4.098E + 4	1.547E + 6	2.466E + 6	2.918E + 6
92	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.190E + 6	2.159E + 6	2.675E + 6
94	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.060E + 5	1.841E + 6	2.418E + 6
96	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	4.703E + 5	1.592E + 6	2.214E + 6
98	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.925E + 5	1.359E + 6	2.032E + 6
100	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	2.163E + 4	1.089E + 6	1.822E + 6
102	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.589E + 5	1.630E + 6
104	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.587E + 5	1.397E + 6

(continued)

TABLE C-7 (continued).

Energy (keV)	50 kVp ^a 2.54 mm	60 kVp ^b 3.21 mm	70 kVp 4.18 mm	80 kVp 4.82 mm	90 kVp 5.41 mm	100 kVp 5.94 mm	110 kVp 6.43 mm	120 kVp 6.89 mm
106	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	3.185E + 5	1.199E + 6
108	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.276E + 5	9.837E + 5
110	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.499E + 4	7.804E + 5
112	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	5.870E + 5
114	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	4.083E + 5
116	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	2.193E + 5
118	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	8.899E + 4
120	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	0.000E + 0	1.182E + 4

^aEach spectrum is normalized to 1 R and individual table entries are in the units of photons/mm² per 2 keV energy interval.

^bSpectra for 50 and 60 kVp are with 1.5 mm added Al filtration.

Appendix D

RADIOPHARMACEUTICAL CHARACTERISTICS AND DOSIMETRY

TABLE D-1. ROUTE OF ADMINISTRATION, LOCALIZATION, CLINICAL UTILITY, AND OTHER CHARACTERISTICS OF COMMONLY USED RADIOPHARMACEUTICALS

Radiopharmaceutical	Method of administration	Delay before imaging	Method of localization	Clinical utility	Comments
⁵⁷ Co Cyanocobalamin (DNI)	ORL	NA <i>In vitro</i> measurement of radioactivity in urine.	Complexed to intrinsic factor and absorbed by distal ileum.	Diagnosis of pernicious anemia and intestinal absorption deficiencies.	Wait ≥24 hr after more than 1 mg IV/IM vitamin B ₁₂ . Have patient fast ≥8 hr before test.
⁵¹ Cr Sodium chromate RBC's (DNI)	IV	NA 20 mL blood sample at 24 hr and every 2-3 days for ≥30 days	Cr ⁶⁺ attaches to hemoglobin: Cr-RBC is a blood pool marker.	Most commonly used for RBC mass and survival. Also used for splenic sequestration studies. Nonimaging <i>in vitro</i> assay via counting serial blood samples.	RBC labeling efficiency 80-90%. Free Cr ³⁺ rapidly eliminated in urine. Splenic sequestration study. ~111 MBq (~3 mCi) of heat-denatured RBC.
¹⁸ F Fluorodeoxyglucose (FDG) (DI)	IV	45 min	Glucose analogue, increased uptake correlates with increased glucose metabolism. Crosses blood-brain barrier metabolized by brain cells FDG is phosphorylated to FDG 6-phosphate trapped in brain tissue for several hours.	Major clinical application in oncology imaging. Can differentiate between recurring tumor and necrotic tissue. Also used for the localization of epileptogenic focus and assessment and localization of brain hypometabolism. Cardiac applications in determining metabolic activity and viability of myocardium.	Positron emission tomography (PET) radiopharmaceutical. Transmission attenuation correction with Ge-68/Ga-68 sources. 511 keV annihilation photons require substantial shielding of source vials and care to prevent crosstalk between other imaging devices.
⁶⁷ Ga Citrate (DI)	IV	Typically 24-72 hr; range, 6-120 hr	Exact mechanism unknown. Accumulates in lysosomes and is bound to transferrin in blood.	Tumor, abscess imaging; Hodgkin's disease; lymphoma, bronchogenic carcinoma, and acute inflammatory lesions.	⁶⁷ Ga uptake influenced by vascularity, increased permeability of tumor cells, rapid proliferation, and low pH. ⁶⁷ Ga secreted in large intestine; bowel activity interferes with abdominal imaging.
¹²³ I Sodium iodide (DI, DNI)	ORL	4-6 hr	Rapidly absorbed from gastrointestinal tract. Trapped and organically bound by thyroid.	Evaluation of thyroid function and/or morphology. Uptake determined by NaI (TI) thyroid probe.	Dose range 3.7 MBq (100 µCi) (uptake study only) to 14.8 MBq (400 µCi) (uptake and imaging). Administered as a capsule.

¹²⁵ I Albumin (DNI)	IV	N/A ~20 mL blood withdrawn ~10 min after dose administration.	Blood pool marker.	Blood and plasma volume determinations.	Dilution principle and hematocrit used to determine blood and plasma volume.
¹³¹ I Sodium iodide (DI, DNI, T)	ORL	24 hr	Rapidly absorbed from gastrointestinal tract, trapped and organically bound by thyroid.	~370 kBq (~10 µCi) used for uptake studies (DNI); ~74–185 MBq (~2–5 mCi) for thyroid carcinoma metastatic survey; ~185–555 MBq (~5–15 mCi) for hyperthyroidism; ~1.1–7.4 GBq (~30–200 mCi) for thyroid carcinoma.	Doses should be manipulated in 100% exhaust fume hood. Iodine is volatile at low pH. 90% of local radiation dose is from beta; 10% gamma.
¹³¹ I Tositumomab, also known as ¹³¹ I anti-B ₁ antibody (T)	IV infusion on 2 treatment days	Whole body counts are obtained on days 1–3 to determine residence time and calculate therapeutic dose.	This agent is a radiolabeled murine IgG _{2a} monoclonal antibody directed against the CD-20 antigen on B cells.	Radioimmunotherapeutic monoclonal antibody for the treatment of chemotherapy-refractory low-grade or transformed low-grade non-Hodgkin's lymphoma.	Cytotoxic effect is the result of both direct antitumor effects of the antibody and the radiation dose delivered by ¹³¹ I. Saturated solution of potassium iodide (SSKI) is administered to block the thyroid. Can typically be administered on an outpatient basis.
¹¹¹ In Capromab pendetide, also known as ProstaScint (DI)	IV	72–120 hr	This agent is a monoclonal antibody to a membrane-specific antigen expressed in many primary and metastatic prostate cancer lesions.	Indicated for patients with biopsy-proven prostate cancer that is thought to be clinically localized after standard diagnostic imaging tests. Also indicated for postprostatectomy patients with high clinical suspicion of occult metastatic disease.	High false-positive rate. Patient management should not be based on ProstaScint results without confirmatory studies. May induce human anti-mouse antibodies, which may interfere with some immunoassays (e.g., prostate-specific antigen).
¹¹¹ In Pentetreotide, also known as Octreoscan (DI)	IV	24–48 hr	Pentetreotide is a diethylenetriamine pentaacetic acid (DTPA) conjugate of octreotide, a long-acting analog of the human hormone somatostatin. Pentetreotide binds to somatostatin receptors on the cell surfaces throughout the body	Localization of primary and metastatic neuroendocrine tumors bearing the somatostatin receptor.	Incubate ¹¹¹ In Pentetreotide at 25°C for at least 30 min. Do not administer in total parenteral nutrition admixtures or lines, because complex glycosylated conjugates may form. Patients should be well hydrated before dose administration.

(continued)

TABLE D-1 (continued).

Radiopharmaceutical	Method of administration	Delay before imaging	Method of localization	Clinical utility	Comments
^{111}In Satumomab pentetide, also known as OncoScint (DI)	IV over a 5-min period	48-72 hr	This agent is a conjugate produced from the murine monoclonal antibody directed to a tumor-associated antigen (TAG-72), a high-molecular-weight glycoprotein expressed differentially by adenocarcinomas.	Colorectal and ovarian carcinoma.	High positive predicted value but low negative predictive value, so negative scans should not be used to guide clinical practice.
^{111}In WBCs (DI)	IV	4-24 hr	^{111}In oxyquinoline is a neutral lipid-soluble complex that can penetrate cell membranes, where the ^{111}In translocates to cytoplasmic components.	Detection of occult inflammatory lesions not visualized by other modalities. Careful isolation and labeling of WBCs is essential to maintain cell viability labeling efficiency ~80%.	^{111}In WBC prepared <i>in vitro</i> with ^{111}In oxyquinoline (In-Oxine) and isolated plasma-free suspension of autologous WBCs. ~30% dose in spleen, ~30% liver, ~5% lung (clears in ~4 hr). ~20% in circulation after injection.
$^{81\text{m}}\text{Kr}$ Krypton gas (DI)	INH	None	No equilibrium or washout phase, so $^{81\text{m}}\text{Kr}$ distribution reflects only regional ventilation.	Lung ventilation studies; however, availability of $^{81\text{Rb}}/^{81\text{m}}\text{Kr}$ generator is limited due to short parent half-life (4.6 hr).	Due to ultrashort half-life (13 sec), only wash-in images can be obtained; however, repeat studies and multiple views can be obtained; radiation dose and hazard are minimal.
$^{32\text{p}}$ Chromic phosphate (T)	IC	NA	Intraperitoneal (IP) or intrapleural (IPL) instillation.	Instilled directly into the body cavity containing a malignant effusion. Dose range, IP, 370-740 MBq (10-20 mCi); IPL, 222-444 MBq (6-12 mCi).	Bluish-green colloidal suspension. Never given IV. Do not use in the presence of ulcerative tumors.
$^{32\text{p}}$ Chromic phosphate (T)	IV	NA	Concentration in blood cell precursors.	Treatment of polycythemia vera (PCV) most common; also myelocytic leukemia, chronic lymphocytic leukemia (CLL).	Clear colorless solution, typical dose range (PCV) 37-296 MBq (1-8 mCi). Do not use sequentially with chemotherapy. WBC should be $\geq 5000/\text{mm}^3$; platelets should be $\geq 150,000/\text{mm}^3$.

¹⁵³ Sr Lexidronam, also known as Quadramet (T)	IV over a 1–2 min period	103 keV photon can be imaged, but this is not essential.	This therapeutic agent is formulated as a tetraphosphonate chelate (EDTMP) with an affinity for areas of high bone turnover associated with the hydroxyapatite (i.e., primary and metastatic bone tumors).	Therapy only. Indicated for the relief of bone pain in patients with painful skeletal metastases.	Patient should be well hydrated. Same precaution as with ⁹⁰ Sr.
⁸⁹ Sr Chloride, also known as Metastron (T)	IV over a 1–2 min period	NA	Calcium analog selectively localizing in bone mineral. Preferential absorption occurs in sites of active osteogenesis (i.e., primary and metastatic bone tumors).	Therapy only. Indicated for the relief of bone pain in patients with painful skeletal metastasis.	Bone marrow toxicity is expected. Peripheral blood cell counts should be monitored at least once every other week. Pain relief expected 7–21 days after dose, so Metastron is not recommended for patients with very short life expectancy.
^{99m} Tc Apcitide, also known as AcuTect (DI)	IV	10–60 min	Binds to the GPIIb/IIIa ($\alpha_2\beta_3$) adhesion-molecule receptors (of the integrin family) found on activated platelets.	Imaging agent for detection of acute venous thrombosis (AVT) in the lower extremities in patients who have signs and symptoms of AVT.	Incubate AcuTect in a boiling water bath for 15 min after adding ^{99m} Tc. Approximately 50% of dose excreted in the urine 2 hr after injection.
^{99m} Tc Depreotide, also known as NeoTect (DI)	IV	2–4 hr	Based on a synthetic peptide with a high-affinity binding to somatostatin receptors in normal and abnormal tissues.	Identifies somatostatin receptor-bearing pulmonary masses. Can distinguish between benign and malignant pulmonary nodules.	Incubate NeoTect in a boiling water bath for 15 min after adding ^{99m} Tc. Do not inject into total parenteral nutrition admixtures or lines because NeoTect may form complex glycosyldepreotide conjugates.
^{99m} Tc Disofenin, also known as HIDA (iminodiacetic acid) (DI)	IV	~5 min serial images every 5 min for ~30 min; and at 40 and 60 min.	Excreted through the hepatobiliary tract into the intestine. Clearance mechanism: hepatocyte uptake, binding, and storage. Excretion into canaliculi via active transport.	Hepatobiliary imaging agent, hepatic duct and gallbladder visualization occurs ~20 min after administration as liver activity decreases. Patient should be fasting >4 hr. Delayed images are obtained at ~4 hr for acute/chronic cholecystitis.	Biliary system not well visualized when serum bilirubin values are higher than ~8 mg/dL; do not use if >10 mg/dL. Normal study: visualized common duct and intestinal tract by ~30 min.

(continued)

TABLE D-1 (continued).

Radiopharmaceutical	Method of administration	Delay before imaging	Method of localization	Clinical utility	Comments
^{99m} Tc DMSA (dimercaptosuccinic acid), also known as Succimer (DI)	IV	1-2 hr	After IV administration ^{99m} Tc Succimer is bound to plasma proteins. Cleared from plasma with a half-life of about 60 min and concentrates in the renal cortex.	Evaluation of renal parenchymal disorders.	Should be administered within 10 min to 4 hr after preparation.
^{99m} Tc Exametazime, also known as Ceretec and HMPAO (DI)	IV	~30-300 min	Rapidly cleared from blood. Lipophilic complex, brain uptake ~4% of injected dose occurs ~1 min after injection.	Detection of altered regional cerebral perfusion in stroke, Alzheimer's disease, seizures, etc. When formulated without methylene blue stabilization, Ceretec can be used to radiolabel leukocytes as an adjunct in the localization of intra-abdominal infection and inflammatory bowel disease.	Administer within 30 min after preparation. Single photon emission computed tomography (SPECT) images only. As with ¹¹¹ In WBCs, careful isolation and labeling of WBCs is essential to maintain cell viability.
^{99m} Tc Macro aggregated albumin (MAA) (DI)	IV	None	~85% ^{99m} Tc MAA mechanically trapped in arterioles and capillaries. Particle size (~15-80 μm). Distribution in lungs is a function of regional pulmonary blood flow.	Scintigraphic evaluation of pulmonary perfusion; ~0.6 million (range 0.2-1.2 million) particles are injected with an adult dose. Not to be used for patients with severe pulmonary hypertension or history of hypersensitivity to human serum albumin.	Mix reagent thoroughly before withdrawing dose; avoid drawing blood into syringe or injecting through IV tubing to prevent nonuniform distribution of particles. Inject patients supine.
^{99m} Tc Medronate, also known as ^{99m} Tc methylenediphosphonate (MDP) (DI)	IV	2-4 hr	Specific affinity for areas of altered osteogenesis. Uptake related to osteogenic activity and skeletal perfusion. ~50% bone, ~50% bladder in 24 hr.	Skeletal imaging agent. Three phase (flow, blood pool, bone uptake); used to distinguish between cellulitis and osteomyelitis. Theory of bone uptake: (a) hydroxyapatite crystal binding; (b) collagen dependent uptake in organic bone matrix.	Normal increase in activity in distal aspect of long bones compared with diaphyses. Higher target to background than pyrophosphate; 85% urinary excretion with MDP vs. 65% with pyrophosphate in 24 hr.

^{99m} Tc Mertiatide, also known as MAG3 (DI)	IV	None	Reversible plasma protein binding. Excreted by kidneys via active tubular secretion and glomerular filtration.	Renal imaging agent. Diagnostic aid for assessment of renal function; split function; renal angiograms; renogram curves, whole kidney and renal cortex. A ^{99m} Tc substitute for ¹³¹ I Hippuran.	~90% of dose excreted in 3 hr. Preparation involves injection of ~2 mL of air in reaction vial and 10 min incubation in boiling water bath within 5 min after ^{99m} Tc is added. Kits are light sensitive and must be stored in the dark.
^{99m} Tc Bicisate, also known as Neurolite (DI)	IV	30–60 min	This is a lipophilic complex that crosses the blood-brain barrier and intact cells by passive diffusion. Cells metabolize this agent to a polar (i.e., less diffusible) compound.	SPECT imaging with Neurolite used as an adjunct to computed tomography and magnetic resonance imaging in the localization of stroke in patients in whom stroke has been diagnosed.	Localization depends on both perfusion and regional uptake by brain cells. Use caution in patients with hepatic or renal impairment.
^{99m} Tc Pentetate, also known as ^{99m} Tc DTPA (DI)	IV	None	Cleared by glomerular filtration; little to no binding to renal parenchyma.	Kidney imaging; assess renal perfusion; estimate glomerular filtration rate (GFR). Historically used as a brain imaging agent to identify excessive vascularity or altered blood-brain barrier.	~5% bound to serum proteins, which leads to a GFR lower than that determined by inulin clearance.
^{99m} Tc Pyrophosphate (DI)	IV	1–6 hr bone imaging; 1–2 hr myocardial imaging	Specific affinity for areas of altered osteogenesis and injured/infarcted myocardium. 50% (bone) and ~50% (bladder) in 24 hr.	Skeletal imaging agent used to demonstrate areas of hyper- or hypo-osteogenesis and/or osseous blood flow; cardiac agent as an adjunct in diagnosis of acute myocardial infarction.	Pyrophosphate also has an affinity for RBCs. Administered IV ~30 min before ^{99m} Tc for <i>in vivo</i> RBC cell labeling. ~75% of activity remains in blood pool.
^{99m} Tc RBCs (DI)	IV	None	Blood pool marker; ~25% excreted in 24 hr. 95% of blood pool activity bound to RBC at 24 hr.	Blood pool imaging including cardiac first-pass and gated equilibrium imaging and detection of sites of gastrointestinal bleeding. Heat-damaged RBCs are used for diagnosis of splenosis and accessory spleen.	<i>In vitro</i> cell labeling; administered within 30 min, labeling efficiency ~95%. Administered with large (19–21 gauge) needle.

(continued)

TABLE D-1 (continued).

Radiopharmaceutical	Method of administration	Delay before imaging	Method of localization	Clinical utility	Comments
^{99m}Tc Sestamibi also known as Cardiolite (DI)	IV	30–60 min	Accumulates in viable myocardium in a manner analogous to ^{201}Tl chloride. Major clearance pathway is the hepatobiliary system.	Diagnosis of myocardial ischemia and infarct. Also used for tumor imaging and thyroid suppression imaging.	Patient should not have fatty foods (e.g., milk) after injection. Low-dose/high-dose regime (typical 3:1 activity ratio) allows same day rest/stress images to be obtained.
^{99m}Tc Sodium pertechnetate (DI)	IV	None for angio/venography; other applications, ~0.5–1 hr.	Pertechnetate ion distributes similarly to the iodide ion; it is trapped but not organified by the thyroid gland.	Primary agent for radiopharmaceutical preparations; thyroid imaging; salivary gland imaging; placental localization; detection of vesicourethral reflux; radionuclide angiography/venography.	Eluate should not be used >12 hr after elution. ^{99m}Tc produced as daughter of Mo/Tc generator.
^{99m}Tc Sulfur colloid (DI)	IV	~20 min	Rapidly cleared by reticuloendothelial (RE) system; ~85% of colloidal particles are phagocytized by Kupffer cells of liver; ~10% in RE system of spleen; ~5% in bone marrow.	Relative functional assessment of liver, spleen, bone marrow, RE system; also used for gastroesophageal reflux imaging; esophageal transit time after oral administration. Patency evaluation of peritoneovenous (LeVeen) shunt, administered in shunt.	Preparation requires addition of acidic solution, 5 min in boiling water bath followed by addition of a buffer to bring pH to 6–7. Plasma clearance half-time ~2–5 min. Larger particles (~100 nm) accumulate in liver and spleen; small particles (<20 nm) in bone marrow.
^{99m}Tc Tetrofosmin, also known as Myoview (DI)	IV	15 min stress, 30 min rest.	Lipophilic cationic complex whose uptake in the myocardium reaches a maximum of ~1.2% of the injected dose at 5 min after injection.	Useful in the delineation of regions of reversible myocardial ischemia in the presence or absence of infarcted myocardium.	Diagnostic efficacy for ischemia and infarct are similar to ^{201}Tl .

²⁰¹ Tl Thallous chloride (DI)	IV	Injected at maximal stress. Image ~10–15 min later. Redistribution image ~4 hr after stress image.	Accumulates in viable myocardium, analogous to potassium.	Identification and semiquantitative assessment of myocardial ischemia and infarction. Also localizes in parathyroid adenomas.	^{99m} Tc-based myocardial perfusion agents are replacing ²⁰¹ Tl due to superior imaging characteristics.
¹³³ Xe Xenon gas (DI)	INH	None	Inhalation; ¹³³ Xe freely exchanges between blood and tissue; concentrates more in fat than in other tissues.	Pulmonary function. Gas is administered by inhalation from a closed respirator system.	Physiologically inert noble gas. Most of the gas that enters circulation returns to lungs and is exhaled. Exhaled ¹³³ Xe gas is trapped (adsorbed) by charcoal filter system.

Note: Radiopharmaceutical used for: D, diagnostic imaging; DNI, diagnostic nonimaging; T, therapy. Method of administration: IV, intravenous; IM, intramuscular; IC, intracavitary; ORL, oral; INH, inhalation. RBC, red blood cell; WBC, white blood cell; NA, not applicable.

TABLE D-2. TYPICAL ADMINISTERED ADULT ACTIVITY, HIGHEST ORGAN DOSE, GONADAL DOSE, AND ADULT EFFECTIVE DOSE FOR COMMONLY USED RADIOPHARMACEUTICALS

Radiopharmaceutical	Typical adult dose			Organ receiving highest dose		Organ dose		Gonadal dose		Effective dose		
	MBq	mCi	Organ	mGy	rad	mGy	rad	mGy	rad	mSv	rem	Source
⁵⁷ Co Cyanocobalamin (IV, no carrier)	0.037	0.001	Liver	1.89	0.189	0.063	ov	0.036	ts	0.16	0.016	1,2
⁵¹ Cr Sodium chromate RBC's	5.6	0.15	Spleen	8.96	0.896	0.46	ov	0.35	ts	0.95	0.095	1,2
¹⁸ F Fluorodeoxyglucose	370	10	Bladder wall	59.20	5.920	5.55	ov	4.44	ts	7.03	0.703	1
⁶⁷ Ga Citrate	185	5	Bone surfaces	116.55	11.655	15.17	ov	10.36	ts	18.50	1.850	1
¹²³ I Sodium iodide (35% uptake)	14.8	0.40	Thyroid	66.60	6.660	0.16	ov	0.07	ts	3.26	0.326	1,2
¹²⁵ I Albumin	0.74	0.02	Heart wall	0.51	0.051	0.15	ov	0.12	ts	0.16	0.016	1,2
¹³¹ I Sodium iodide (35% uptake)	3700	100	Thyroid	1.85E+06	1.85E+05	155.40	ov	96.20	ts	NA	NA	1,2
¹¹¹ In Pentetreotide, also known as Octreoscan	222	6	Spleen	126.54	12.654	5.99	ov	3.77	ts	11.99	1.199	1
¹¹¹ In WBCs	18.5	0.50	Spleen	101.8	10.18	2.22	ov	0.83	ts	6.66	0.666	1,2
^{81m} Kr Krypton gas	370	10	Lung	0.08	0.008	6.3E-05	ov	6.3E-06	ts	0.01	0.001	1,2
³² P Chromic phosphate	148	4	Red marrow	1628.00	162.800	109.52	ov	109.52	ts	NA*	NA*	2
¹⁵³ Sm Lexidronam, also known as Quadramet	2590	70	Bone surfaces	1.75E+04	1.75E+03	22.27	ov	13.99	ts	NA*	NA*	4
⁸⁹ Sr Chloride, also known as Metastron	148	4	Bone surfaces	2516.00	251.600	118.40	ov	118.40	ts	NA	NA	5
^{99m} Tc Acpitide, also known as AcuTect	740	20	Bladder wall	44.40	4.440	4.66	ov	3.92	ts	6.8	0.68	3,4
^{99m} Tc Depreotide, also known as NeoTect	740	20	Kidneys	66.60	6.660	3.11	ov	22.94	ts	11.20	1.12	3,4
^{99m} Tc Disofenin, also known as HIDA (iminodiacetic acid)	185	5	Gall bladder wall	20.35	2.035	3.52	ov	0.28	ts	3.15	0.315	1
^{99m} Tc DMSA (dimercaptosuccinic acid) also known as Succimer	185	5	Kidneys	33.30	3.330	0.65	ov	0.33	ts	1.63	0.163	1
^{99m} Tc Exametazime, also known as Ceretec and HMPAO	740	20	Kidneys	25.16	2.516	4.88	ov	1.78	ts	6.88	0.688	1

^{99m} Tc Macroaggregated albumin (MAA)	148	4	Lungs	9.77	0.977	0.27 0.16	ov ts	1.63	0.163	1
^{99m} Tc Medronate also known as ^{99m} Tc methylenediphosphonate (MDP)	740	20	Bone surfaces	46.62	4.662	2.66 1.78	ov ts	4.22	0.422	1
^{99m} Tc Mertiatide also known as MAG3 (Normal renal function)	185	5	Bladder wall	20.35	2.035	1.00 0.685	ov ts	1.30	0.130	1
^{99m} Tc Bicisate, also known as ECD and Neurolite	740	20	Bladder wall	53.58	5.358	3.61 2.45	ov ts	6.73	0.673	3
^{99m} Tc Pentetate, also known as ^{99m} Tc DTPA	370	10	Bladder wall	22.94	2.294	1.55 1.07	ov ts	8.14	0.814	1
^{99m} Tc Pyrophosphate	555	15	Bone surfaces	34.97	3.497	2.00 1.33	ov ts	3.16	0.316	1
^{99m} Tc RBCs	740	20	Heart	17.02	1.702	2.74 1.70	ov ts	5.18	0.518	1
^{99m} Tc Sestamibi, also known as Cardiolite	740	20	Gall bladder wall	28.86	2.886	6.73 2.81	ov ts	6.66	0.666	1
^{99m} Tc Sestamibi, also known as Cardiolite (Stress)	296	8	Gall bladder wall	9.77	0.977	2.40 1.10	ov ts	2.34	0.234	1
^{99m} Tc Sodium pertechnetate	370	10	Upper large intestine	21.09	2.109	3.70 1.04	ov ts	4.81	0.481	1
^{99m} Tc Sulfur colloid	296	8	Spleen	22.20	2.220	0.65 0.17	ov ts	2.78	0.278	1
^{99m} Tc Tetrofosmin, also known as Myoview (Rest)	740	20	Gall bladder wall	26.64	2.664	6.22 1.78	ov ts	5.62	0.562	1
^{99m} Tc Tetrofosmin, also known as Myoview (Stress)	296	8	Gall bladder wall	7.99	0.799	2.25 0.86	ov ts	2.07	0.207	1
²⁰¹ Tl Thallous chloride	74	2	Thyroid	45.9	4.59	7.4 14.8	ov ts	11.8	1.18	5
¹³³ Xe Xenon gas (rebreathing for 5 min)	555	15	Lungs	0.61	0.061	0.41 0.38	ov ts	0.41	0.041	1,2

Note: NA, not applicable; ov, ovary; RBC, red blood cell; ts, testis; WBC, white blood cell.

^aThe effective dose is not a relevant quantity for therapeutic doses of radionuclides because it relates to stochastic risk (e.g., cancer and genetic effects) and is not relevant to deterministic (i.e., non-stochastic) risks (e.g., acute radiation syndrome).

Sources: (1) Annals of the International Commission on Radiological Protection. *Publication 80: Radiation dose to patients from radiopharmaceuticals*. Tarrytown, NY: Elsevier Science Inc., 1999. (Consult this reference for dosimetry information on other radiopharmaceuticals.)

(2) The International Commission on Radiological Protection. *Publication 53: Radiation dose to patients from*

radiopharmaceuticals. Elmsford, NY: Pergamon Press, 1988. (Consult this reference for dosimetry information on other radiopharmaceuticals.)

(3) Courtesy of Richard B. Sparks, PhD, of CDE, Inc. Dosimetry Services, Knoxville, TN.

(4) Information supplied by the manufacturer.

(5) Stabin MG, Stubbs JB, and Toohey RE. *Radiation dose estimates for radiopharmaceuticals*, NUREG/CR-6345. Washington DC: Department of Energy and the Department of Health and Human Services, 1996.

TABLE D-3. EFFECTIVE DOSE PER UNIT ACTIVITY ADMINISTERED TO PATIENTS AGE 15, 10, 5, AND 1 YEAR FOR COMMONLY USED DIAGNOSTIC RADIOPHARMACEUTICALS

Radiopharmaceutical	Effective dose							
	15 yr		10 yr		5 yr		1 yr	
	mSv/MBq	rem/mCi	mSv/MBq	rem/mCi	mSv/MBq	rem/mCi	mSv/MBq	rem/mCi
¹⁸ F Fluorodeoxyglucose	0.025	0.093	0.036	0.133	0.050	0.185	0.095	0.352
⁶⁷ Ga Citrate	0.130	0.481	0.200	0.740	0.330	1.221	0.640	2.368
¹²³ I Sodium iodide (35% uptake)	0.405	1.500	0.595	2.200	1.270	4.700	2.378	8.800
¹¹¹ In Pentatreotide, also known as Octreotide	0.071	0.263	0.100	0.370	0.160	0.592	0.280	1.036
¹¹¹ In WBCs	0.459	1.700	0.703	2.600	1.054	3.900	1.919	7.100
^{99m} Tc Disofenin, also known as HIDA (iminodiacetic acid)	0.021	0.078	0.029	0.107	0.045	0.167	0.100	0.370
^{99m} Tc DMSA (dimercaptosuccinic acid) also known as Succimer	0.011	0.041	0.015	0.056	0.021	0.078	0.037	0.137
^{99m} Tc Exametazime, also known as Ceretec and HMPAO	0.011	0.041	0.017	0.063	0.027	0.100	0.049	0.181
^{99m} Tc Macro aggregated albumin (MAA)	0.016	0.059	0.023	0.085	0.034	0.126	0.063	0.233
^{99m} Tc Medronate, also known as ^{99m} Tc Methylenediphosphonate (MDP)	0.007	0.026	0.011	0.041	0.014	0.052	0.027	0.100
^{99m} Tc Mertiatide also known as MAG3	0.009	0.033	0.012	0.044	0.012	0.044	0.022	0.081
^{99m} Tc Bicisate, also known as ECD and Neulolite	0.012	0.043	0.012	0.043	0.018	0.067	0.026	0.096
^{99m} Tc Pentetate, also known as ^{99m} Tc DTPA	0.006	0.022	0.008	0.030	0.009	0.033	0.016	0.059
^{99m} Tc Pyrophosphate	0.007	0.026	0.011	0.041	0.014	0.052	0.027	0.100
^{99m} Tc RBCs	0.009	0.033	0.014	0.052	0.021	0.078	0.039	0.144
^{99m} Tc Sestamibi, also known as Cardiolite (Rest)	0.012	0.044	0.018	0.067	0.028	0.104	0.053	0.196
^{99m} Tc Sestamibi, also known as Cardiolite (Stress)	0.010	0.037	0.016	0.059	0.023	0.085	0.045	0.167
^{99m} Tc Sodium pertechnetate	0.017	0.063	0.026	0.096	0.042	0.155	0.079	0.292
^{99m} Tc Sulfur colloid	0.012	0.044	0.018	0.067	0.028	0.104	0.050	0.185
^{99m} Tc Tetrofosmin, also known as Myoview (Rest)	0.010	0.037	0.013	0.048	0.022	0.081	0.043	0.159
^{99m} Tc Tetrofosmin, also known as Myoview (Stress)	0.008	0.031	0.012	0.044	0.018	0.067	0.035	0.130
²⁰¹ Tl Thallous chloride	0.300	1.110	1.200	4.440	1.700	6.290	2.800	10.360

Note: RBC, red blood cell; WBC, white blood cell.

Sources: (1) Annals of the International Commission on Radiological Protection. *Publication 80: Radiation dose to patients from radiopharmaceuticals*. Tarrytown, NY: Elsevier Science Inc., 1999.
(2) Courtesy of Richard B. Sparks, PhD, of CDE, Inc. Dosimetry Services, Knoxville, TN.

TABLE D-4. ABSORBED DOSE ESTIMATES TO THE EMBRYO/FETUS PER UNIT ACTIVITY ADMINISTERED TO THE MOTHER FOR COMMONLY USED RADIOPHARMACEUTICALS

Radiopharmaceutical	Dose at different stages of gestation							
	Early		3 mo		6 mo		9 mo	
	mGy/MBq	rad/mCi	mGy/MBq	rad/mCi	mGy/MBq	rad/mCi	mGy/MBq	rad/mCi
⁵⁷ Co Cyanocobalamin (normal w/o flushing dose)	1.5000	5.5500	1.0000	3.7000	1.2000	4.4400	1.3000	4.8100
¹⁸ F Fluorodeoxyglucose	0.0270	0.0999	0.0170	0.0629	0.0094	0.0348	0.0081	0.0300
⁶⁷ Ga Citrate	0.0930	0.3441	0.2000	0.7400	0.1800	0.6660	0.1300	0.4810
¹²³ I Sodium iodide (25% maternal thyroid uptake)	0.0200	0.0740	0.0140	0.0518	0.0110	0.0407	0.0098	0.0363
¹²⁵ I Albumin	0.2500	0.9250	0.0780	0.2886	0.0380	0.1406	0.0260	0.0962
¹³¹ I Sodium iodide (25% maternal thyroid uptake)	0.0720	0.2664	0.0680	0.2516	0.2300	0.8510	0.2700	0.9990
¹¹¹ In Pentetreotide, also known as Octreotide	0.0820	0.3034	0.0600	0.2220	0.0350	0.1295	0.0310	0.1147
¹¹¹ In WBCs	0.1300	0.4810	0.0960	0.3552	0.0960	0.3552	0.0940	0.3478
^{99m} Tc Disofenin, also known as HIDA (iminodiacetic acid)	0.0170	0.0629	0.0150	0.0555	0.0120	0.0444	0.0067	0.0248
^{99m} Tc DMSA (dimercaptosuccinic acid), also known as Succimer	0.0051	0.0189	0.0047	0.0174	0.0040	0.0148	0.0034	0.0126
^{99m} Tc Exametazime, also known as Cerete and HMPAO	0.0087	0.0322	0.0067	0.0248	0.0048	0.0178	0.0036	0.0133
^{99m} Tc Macro aggregated albumin (MAA)	0.0028	0.0104	0.0040	0.0148	0.0050	0.0185	0.0040	0.0148
^{99m} Tc Medronate, also known as Methylenediphosphonate (MDP)	0.0061	0.0226	0.0054	0.0200	0.0027	0.0100	0.0024	0.0089
^{99m} Tc Meritattide, also known as MAG3	0.0180	0.0666	0.0140	0.0518	0.0055	0.0204	0.0052	0.0192
^{99m} Tc Bicisate, also known as ECD and Neurolite	0.013	0.0481	0.0100	0.0370	0.0055	0.0200	0.0044	0.0163
^{99m} Tc Pentetate also known as ^{99m} Tc DTPA	0.0120	0.0444	0.0087	0.0322	0.0041	0.0152	0.0047	0.0174
^{99m} Tc Pyrophosphate	0.0060	0.0222	0.0066	0.0244	0.0036	0.0133	0.0029	0.0107
^{99m} Tc RBCs	0.0064	0.0237	0.0043	0.0159	0.0033	0.0122	0.0027	0.0100
^{99m} Tc Sestamibi, also known as Cardiolite (Rest)	0.0150	0.0555	0.0120	0.0444	0.0084	0.0311	0.0054	0.0200
^{99m} Tc Sestamibi, also known as Cardiolite (Stress)	0.0120	0.0444	0.0095	0.0352	0.0069	0.0255	0.0044	0.0163
^{99m} Tc Sodium pertechnetate	0.0110	0.0407	0.0220	0.0814	0.0140	0.0518	0.0093	0.0344
^{99m} Tc Sulfur colloid	0.0018	0.0067	0.0021	0.0078	0.0032	0.0118	0.0037	0.0137
^{99m} Tc Teboroxime, also known as Cardiotec	0.0089	0.0329	0.0071	0.0263	0.0058	0.0215	0.0037	0.0137
²⁰¹ Tl Thallous chloride (Rest)	0.0970	0.3589	0.0580	0.2146	0.0470	0.1739	0.0270	0.0999
¹³³ Xe Xenon gas (rebreathing 5 min)	0.00041	0.00152	0.00005	0.00018	0.00004	0.00013	0.00003	0.00010

Note: RBC, red blood cell; WBC, white blood cell.

Sources: (1) Stabin MG. *Fetal dose calculation workbook (ORISE 97-0961)*, Radiation Internal Dose. Information Center, Oak Ridge Institute for Science and Education, Oak Ridge, TN, 1997.

(2) Courtesy of Richard B. Sparks, PhD, CDE, Inc., Dosimetry Services, Knoxville, TN.

Appendix E

INTERNET RESOURCES

There are a number of professional societies, governmental organizations, and other entities that provide, on the Internet, relevant educational, reference, and other material related to the physics of medical imaging. Some of these sites are listed here.

Advanced Medical Publishers: www.advmedpub.com
American Association of Physicists in Medicine (AAPM): www.aapm.org
American Board of Radiology (ABR): www.theabr.org
American College of Medical Physics (ACMP): www.acmp.org
American College of Radiology (ACR): www.acr.org
American Institute of Physics (AIP): www.aip.org
American Institute of Ultrasound in Medicine (AIUM): www.aium.org
American Journal of Roentgenology (AJR): www.arrs.org/ajr
American National Standards Institute (ANSI): www.ansi.org
American Roentgen Ray Society (ARRS): www.arrs.org
American Society of Radiologic Technologists (ASRT): www.asrt.org
British Institute of Radiology (BIR): www.bir.org.uk
Canadian College of Physicists in Medicine (CCPM): www.medphys.ca
Conference of Radiation Control Program Directors (CRCPD): www.crcpd.org
Digital Imaging and Communications in Medicine (DICOM): www.xray.hmc.psu.edu/dicom
Electronic Medical Physics World: www.medphysics.wisc.edu/~empw
Food and Drug Administration (FDA): www.fda.gov
Health Physics Society (HPS): www.hps.org
Institute of Electrical and Electronics Engineers: www.ieee.org
The Institute of Physics and Engineering in Medicine (IPEM): www.ipem.org.uk
Intersocietal Commission for the Accreditation of Echocardiography Laboratories (ICAEL): www.icael.org
International Atomic Energy Agency (IAEA): www.iaea.org
International Commission on Non-Ionizing Radiation Protection: www.icnirp.de
International Commission on Radiation Units and Measurements (ICRU): www.icru.org
International Commission on Radiological Protection (ICRP): www.icrp.org
International Organization of Medical Physicists (IOMP): www.iomp.org
Joint Commission for Accreditation of Healthcare Organizations: www.jcaho.org
Medical Physics Journal: www.medphys.org

Medical Physics Publishing: www.medicalphysics.org

National Council on Radiation Protection and Measurements (NCRP):
www.ncrp.com

National Electrical Manufacturer's Association (NEMA): www.nema.org

National Radiological Protection Board (NRPB): www.nrpb.org.uk

National Technical Information Service: www.ntis.gov

Radiation Research Society: www.radres.org

Radiographics Journal: www.rsnaajnl.org

Radiological Society of North America (RSNA): www.rsna.org

Radiology Journal: www.rsnaajnl.org

Society for Computer Applications in Radiology (SCAR): www.scarnet.org

Society of Nuclear Medicine (SNM): www.snm.org

SPIE—The International Society for Optical Engineering (SPIE): www.spie.org

U.S. National Institute of Standards and Technology (NIST): www.nist.gov

U.S. Nuclear Regulatory Commission (NRC): www.nrc.gov



SUBJECT INDEX



دانشکده مهندسی هسته ای
تهیه و تنظیم : نوید حسن پور

Page numbers followed by *f* indicate figures; page numbers followed by *t* indicate tables

A

- ABC. *See* Automatic brightness control
- Absolute risk, radiation biology, 846
- Absolute speed, 161, 161*f*
- Absorption
 - efficiency, 154–155, 154*f*, 155*f*
- Absorbed Dose. *See* dose
- Acceleration, 867
- Access
 - random, 70
 - sequential, 70
- Accuracy, characterization of data, counting statistics, 661
- Acoustic impedance, ultrasound, 14, 477, 477*t*
- Acoustic power, pulsed ultrasound intensity, 549–551, 549*f*, 550*t*
- Acquisition, interface, 77
- Acquisition time, magnetic resonance imaging, 430
- Activator, development, film processor, 180
- Active transport, radiopharmaceuticals, 621–622
- Activity, radionuclide decay, defined, 589, 590*t*
- Acute radiation syndrome (ARS), 831–837
 - gastrointestinal syndrome, 835
 - hematopoietic syndrome, 833–835, 834*f*
 - neurovascular syndrome, 835–836
 - sequence of events, 832–833, 832*f*, 833*t*
- Adaptive histogram equalization, image quality, 315
- ADCs. *See* Analog-to-digital converters
- Added filtration, x-ray, 114
- Advisory bodies, radiation protection, 789–790
- Air gap
 - as alternative to antiscatter grid, 172–173, 173*f*
 - scattered radiation, mammography, 207–210, 209*f*
- Air kerma, 52–53
- ALARA. *See* As low as reasonably achievable
- Aliasing, image quality and, 283–287, 284*f*, 285*f*, 540*f*, 541
- Alpha decay, nuclear transformation, 593
- Alternating (AC) electric current, 116
 - defined, 872–873, 873*f*
- American Standard Code for Information Interchange (ASCII), binary representation of text, 65
- Ampere, defined, 870
- Amplitude, electromagnetic wave, 17
- Analog data
 - conversion to digital form, 68–69
 - representation, 66
- Analog-to-digital converters (ADCs), 68
- Anger camera. *See* Scintillation camera
- Angiography,
 - digital subtraction angiography, 321–322
 - magnetic resonance imaging, 442–446
 - phase contrast, 445, 446*f*
 - time-of-flight, 443–445, 444*f*
- Annihilation coincidence detection (ACID),
 - emission tomography, 719
- Annihilation photons, positron annihilation, 36
- Annihilation radiation, 9. *See* Positron annihilation radiation
- Annual limits on intake (ALIs), 791
- Anode, 87, 105–112, 106*f*
 - cooling chart, 142–143, 142*f*
 - grounded, 195
 - heat input chart, 142–143, 142*f*
 - mammography, 194–195, 195*f*, 196*f*
- Antiscatter grid, 168–173, 168*f*, 207–210
 - air gap as alternative to, 172–173, 208
 - artifacts, 172
 - Bucky factor, 171
 - moving grid, 171
- Aperture blurring, image quality, 286–287, 286*f*
- Array processor, 78, 78*f*
- Artifacts
 - antiscatter grid, 172, 172*f*
 - gradient field, 448
 - machine-dependent, 447
 - magnetic resonance imaging, 447–457
 - processor
 - Hurter and Driffield curve effects, 181–182, 181*f*
 - radiofrequency, 449–450
 - radiofrequency coil, 449
 - susceptibility, 447–448
 - ultrasound, 524–531
 - x-ray computed tomography, 369–372
- As low as reasonably achievable (ALARA), 792

ASCII. *See* American Standard Code for Information Interchange

Atom

- structure, 21–29
 - atomic binding energy, 27–28
 - atomic nucleus, 24–29
 - atomic number, 24
 - binding energy, 21–22
 - electron binding energy, 21–22, 27–28
 - electron cascade, 23
 - electron orbit, 21–22
 - electronic structure, 21–22
 - energy level, 24
 - fission, 28–29
 - ground state, 24
 - K shell, 21–22
 - mass defect, 27–28
 - mass number, 24
 - metastable state, 24
 - nucleon, 24
 - quantum levels, 21–22
 - quantum numbers, electron shell, 21–22
 - valence shell, 22

Atomic binding energy, 27–28

Atomic number, 24

Attenuation

- electromagnetic radiation, 45–52
 - beam hardening, 51–52, 51*f*
 - effective energy, 50, 50*t*
 - half value layer, 48–52
 - homogeneity coefficient, 51
 - linear attenuation coefficient, 45–47
 - mass attenuation coefficient, 47–48
 - mean free path, 50
- ultrasound, 481–483, 481*t*, 482*f*

Auger electron, 23

Automatic brightness control (ABC), fluoroscopy, 246–248, 247*f*

Automatic exposure control (AEC). *See* phototimer system

Autotransformers, x-ray, 120–121, 120*f*

Average glandular dose, radiation dosimetry, mammography, 222–223

Average gradient, defined for screen-film systems, 160, 160*f*

Axial resolution, ultrasound, 497–498, 498*f*

B

Backprojection

- filtered
 - single photon emission computed tomography (SPECT), 707–709
 - x-ray computed tomography, 353–355
- simple, x-ray computed tomography, 351–353, 352*f*

Bandwidth,

- network, 556
- video camera, fluoroscopy, 241

Beam hardening, 51–52, 51*f*

x-ray computed tomography, 369–371, 370*f*, 371*f*

Beta minus decay, nuclear transformation, 594–595, 594*f*

Beta-minus particles, 20

Beta-plus decay, nuclear transformation, 595–596, 596*f*

BGO. *See* Bismuth germanate

Binary representation, 62, 62*t*

Binding energy, 22*f*

electron, 21–22

nuclear, 27–28, 28*f*

Bismuth germanate (BGO), inorganic crystalline scintillators in radiology, 638

Bit, 63, 64*t*

Black level, display of images, 578

Blood oxygen level-dependent (BOLD), 409–410

Blurring, physical mechanisms of spatial resolution, 265, 266*f*

BOLD. *See* Blood oxygen level-dependent

Bone kernels, x-ray computed tomography, 355–356

Bounce point, short tau inversion recovery, magnetic resonance imaging, 401

“Bow-tie” filters, computed tomography, 114

Bragg peak, 32–33

Breast cancer, radiation induced carcinogenesis, 850–851

Breast-feeding, radioiodine therapy, home care, 787

Bremsstrahlung

defined, 35

process, x-ray production, 97–100, 99*f*

radiation, 35–36, 35*f*

Bremsstrahlung spectrum

filtered, 98

unfiltered, 98

Brightness gain, image intensifier, fluoroscopy, 236

Broad-beam geometry, 48–49, 48*f*

Bucky factor, 171, 171*f*

mammography, 210

Byte, 63

C

Calcium tungstate, inorganic crystalline scintillators in radiology, 638

Cancer development, 839

Cardiology catheterization suite, 251

Carrier wave, and image quality, 256

Cassette, in film screen radiography, 148–149, 148*f*

Catalyst, developer, film processing, 177

Cathode, 87, 102–105, 103*f*

mammography, 194

Cathode ray tube (CRT), 77, 87*f*, 87–90, 576–580

- CCD. *See* Charged-coupled devices
- Cellular radiobiology, 818–827
- cell survival curves, 821–822, 822*f*
 - cellular radiosensitivity, 821
 - cellular targets, 818
 - DNA, 818–821, 819*f*, 820*f*
- Center for Devices and Radiological Health (CDRH), 251
- Center of rotation, emission tomography, 715
- Central processing unit, 72–73
- Cesium-137, spectrum of, 650–651, 650*f*
- Chain reactions, 608–609, 608*f*, 610*f*
- Characteristic curve of film, 159–163
- Characteristic x-ray, 23
- Charged-coupled devices (CCD), 297–299, 298*f*, 299*f*
- Chatter, processor artifacts, 183
- Chemical shift artifacts, magnetic resonance imaging, 453–455, 453*f*
- Chemotaxis, radiopharmaceuticals, 622
- Chromatid aberrations, cellular radiobiology, 820
- Chromosome aberrations, cellular radiobiology, 819
- Cine-radiography cameras, 243–244
- Classical scattering. *See* Rayleigh scattering
- Code of Federal Regulations (CFR), 788–789
- Coherent scattering. *See* Rayleigh scattering
- Coil quality factor, magnetic resonance imaging, 441–442
- Collimator
- axial, 725–727, 726*f*, 732–733
 - converging, 676
 - diverging, 676
 - fan-beam, 676, 710–711, 719
 - mammography, 203–204
 - nuclear imaging, 669
 - parallel-hole, 674
 - pinhole, 675
 - scintillation camera, 674–676, 675*f*, 684–686, 685*f*, 686*f*
 - x-ray, 114, 115*f*
- Collimator pitch, x-ray computed tomography, 345–346, 345*f*
- Collimator resolution, performance, scintillation camera, 679
- Committed dose equivalent, radiation protection, 790
- Compartmental localization, radiopharmaceuticals, 620
- Compression of breast, mammography, 207, 208*f*
- Compression, images, 575–576, 575*t*
- Compton scattering, 38–40, 38*f*, 39*f*, 43, 43*f*, 45*f*
- Computed radiography
- photostimulable phosphor system, 293, 294*f*, 295*f*, 296*f*
- Computed tomography (CT), 6. *See* Positron emission tomography; Single photon emission computed tomography; X-ray computed tomography
- Computed tomography dose index (CTDI), 364
- Computer, 70–81
- Computer-aided detection (*also* diagnosis) 86
- Computer networks, 555–565
- Computers in nuclear imaging, 695–702
- Conductor, defined, 873–874
- Conformance statement, DICOM, 571
- Constant potential x-ray generator, 129
- Containment structure, radionuclides, 609
- Contamination control, radiation protection, 781–782
- Continuous fluoroscopy, 244
- Contrast, 13–14
- enhancement
 - computer
 - translation table, 90–91, 90*f*, 91*f*
 - windowing, 91, 92*f*
 - computer display, 86–90
 - and image quality, 255–263, 256*f*
 - ratio, image intensifier, fluoroscopy, 237
 - sensitivity
 - noise, and image quality, ultrasound, 525–526
 - and spatial resolution, magnetic resonance imaging, 439
- Contrast-detail curves, image quality, 287–288, 287*f*
- Contrast reduction factor, scattered radiation, 168
- Contrast resolution
- image quality, fluoroscopy, 248–249
 - vs spatial resolution, digital imaging, 315–316, 315*f*
 - x-ray computed tomography, 369
- Converging collimator, 676
- Conversion
- factor, image intensifier, 236
- Convolution
- and spatial resolution, 312–314, 312*f*, 313*f*
- Counting statistics, 661–668
- data characterization, 661–662
 - estimated standard deviation, 666–668
 - probability distribution functions, for binary processes, 662–666, 663*t*
 - propagation of error, 666–668, 667*t*
 - random error, 661
 - sources of error, 661
 - systematic error, 661
- CPU. *See* Central processing unit
- Critical, nuclear reactor, 608
- Cross-excitation, magnetic resonance imaging, 442
- CRT. *See* Cathode ray tube
- CT number. *See* Hounsfield unit
- Current mode, radiation detection, 628–630
- Cyclotron-produced radionuclide, 603–606, 604*f*, 606*f*

D

- DAC. *See* Digital-to-analog converter
- Data conversion, 66–69, 67*f*
 analog-to-digital conversion, 68–69
 digital-to-analog conversion, 69
 digitization, 68
 electronic noise, 67
 Nyquist limit, 68
 quantization, 68
 sampling, 68
 signal-to-noise ratio, 69
- Daughter nuclide, 27
- Daylight processing, 184
- Decay constant, radionuclide decay defined, 589–590
- Decay equation, radionuclide decay, 591–593, 592*f*
- Decay schemes
 flourine-18, 601–602, 602*f*
 molybdenum-99, 598–600, 600*f*
 nuclear transformation, 598–602, 598*f*
 phosphorous-32, 598, 599*f*
 radon-220, 598, 599*f*
 technetium-99m, 600–601, 601*f*
- Decibel
 scale, and relative intensity values, 476*t*
 sound, 475, 476*t*
- Decimal form, 62, 62*t*
- Decision criterion, receiver operating characteristic curves, 288
- Decision matrix, 2 x 2
 image quality, 289
- Deconvolution, and spatial resolution, 314
- Delta function, and spatial resolution, 313
- Delta ray, particle, 31, 32*f*
- Densitometer, 186
 defined, 158
- Detection efficiency, radiation detection, 630–631, 631*f*
- Detective quantum efficiency
 digital radiography, 308
 image intensification, fluoroscopy, 238, 238*f*
- Detector aperture width, image quality, 283
- Detector pitch
 spatial resolution, x-ray computed tomography, 368
 x-ray computed tomography, 345–346, 345*f*
- Developer. *See* Film processor
- Diamagnetic materials, 879
 defined, 374
- DICOM. *See* Digital Imaging and Communications in Medicine
- Diffusion contrast, magnetic resonance imaging, 409–411
- Digital full-field mammography, 221
- Digital image correction, 309–310, 310*f*
 flat field image, 309
- Digital image processing, 309–315
 digital image correction, 309–310, 310*f*
 global processing, 310–312, 311*f*
- Digital images, acquisition of, 568–571
 film digitizers, 568–571, 569*f*, 570*f*
 image formats, 568
- Digital Imaging and Communications in Medicine (DICOM), 571
- Digital mammography, 304–307, 305*f*, 306*f*
- Digital photo-spot, 234
- Digital radiography
 charged-coupled devices, 297–299, 298*f*, 299*f*
 computed tomography, 293–297, 294*f*, 295*f*, 296*f*
 contrast vs spatial resolution, 315–316
 digital image processing, 309–315
 digital mammography, 304–307, 305*f*, 306*f*
 flat panel detectors, 300–304
 hard copy vs soft copy display, 308
 implementation, 307
 patient dose considerations, 308
 photostimulable phosphor system, 293
 picture archiving and communication system, 293
- Digital subtraction angiography (DSA), temporal subtraction, 321–323, 322*f*
- Digital-to-analog converter, 69
- Digital tomosynthesis, 320–321, 320*f*
- Digitization, 68
- Diode, x-ray, 121–122, 121*f*, 122*f*, 628, 640–644
- Direct current (DC), 116
 defined, 872–873, 873*f*
- Direct detection flat panel system, 303–304, 304*f*
- Disk, data storage
 magnetic, 74–77, 74, 573
 optical, 74–77, 573
- Display, computer, 86–90
 contrast enhancement, 90–91
 digital conversion to analog video signal, 86–87, 86*f*
 hard copy device, 93
- Display of images, 576–582
 ambient viewing conditions, 580–581
 calibration of display workstations, 580
 display workstations, 576–579, 577*f*, 578*t*
 hard copy devices, 581–582
 monitor pixel formats, 579–580
- Display workstations, 576–579, 577*f*, 578*t*
 calibration of, 580
- Displayed contrast, and image quality, 262–263, 262*f*
- Distance, from radiation source, exposure control, 756–758, 757*f*, 758*t*
 inverse square law, 756
- Diverging collimator, 676
- DNA, radiobiology, 818–821, 819*f*, 820*f*
- Doppler frequency shift, 531–533, 532*f*, 533*t*
- Doppler ultrasound, 13, 531–544
 aliasing, 541
 color flow imaging, 541–543, 542*f*

- continuous Doppler operation, 534–535, 534*f*
- Doppler frequency shift, 531–533, 532*f*, 533*t*
- Doppler spectral interpretation, 538–540, 539*f*, 540*f*
- duplex scanning, 537–538
- power Doppler, 543–544, 544*f*
- pulsed Doppler operation, 535–537, 535*f*, 536*f*
- quadrature detection, 535
- Dose
 - absorbed dose, definition, 53
 - cumulated dose, 807
 - dose indices, 797*t*
 - effective dose, definition, 57, 796
 - effective dose equivalent, definition, 58
 - entrance skin exposures (ESESs), 797*t*
 - equivalent dose, definition, 57
 - fetal dose calculation, 802–803, 802*t*
 - fluoroscopy, 251–254
 - patient dose, 251–252, 252*f*
 - personnel dose, 251–252, 253*f*
 - roentgen-area product meters, 253–254
 - interventional procedures, 799*t*
 - mammography, 222–224, 223*t*
 - mean glandular dose, 222
 - radiopharmaceutical, 800, 900*t*–911*t*
 - x-ray computed tomography, 363–365, 363*f*, 364*f*, 365*t*
 - x-ray dosimetry, 800–805, 801*f*
 - x-ray procedures, 800
- Dose calibrator
 - for assaying radiopharmaceuticals, 659–660
 - molybdenum concentration, 660–661
- Dosimeter
 - defined, 628
 - radiation detection, 628
- Dosimetric quantity. *See* Quantities, radiation
- Dosimetry. *See* Personnel dosimetry; Radiation dosimetry; Radiopharmaceuticals dosimetry
- Doubling dose, estimating genetic risk, 852
- Dry processing, 184–185
- Dual-energy subtraction, 323–325, 324*f*, 325*f*
- Duplex scanning, Doppler ultrasound, 537–538
- Dynamic, frame mode acquisition, image acquisition, 696–697
- Dynamic aperture, ultrasound beam, 495
- Dynamic range, Hurter and Driffield curve, 163
- E**
 - Echo planar image acquisition, 435–436
 - Edge spread function (ESF), and spatial resolution, 269
 - Effective dose, 57–58
 - Effective dose equivalent, 58
 - Effective energy, 50
 - Effective focal spot length, 109–110
 - Efficiency
 - conversion efficiency, 153–154, 153*f*, 637
 - detection efficiency, 630–631, 631*f*
 - geometric, 631
 - intrinsic, 631
 - quantum detection efficiency, 154–155
 - radiation detector, 630–631
 - Electrical charge, defined, 868–869
 - Electrical current, 870, 871*f*
 - Electrical dipole, 869
 - Electrical field, 869, 869*f*
 - Electrical force, 869
 - Electrical potential difference, defined, 870–872
 - Electrical power, defined, 872
 - Electromagnetic induction, 116, 117*f*
 - Electromagnetic radiation, 17–19, 19*f*
 - attenuation, 45–52
 - beam hardening, 51–52
 - effective energy, 50
 - half value layer, 48–50
 - homogeneity coefficient, 51
 - linear attenuation coefficients, 45–48
 - mass attenuation coefficients, 47–48
 - mean free path, 50
 - interactions, 37–44
 - Compton scattering, 38–40
 - pair production, 44
 - photoelectric absorption, 40–43
 - Rayleigh scattering, 37–38
 - particle characteristics, 19
 - release of, 32*f*
 - Electromagnetic spectrum, 3, 17, 18*f*
 - Electromagnetic wave characteristics, 17–18, 18*f*
 - amplitude, 17
 - frequency, 17
 - period, 17
 - wavelength, 17
 - Electron
 - anode, target of, 97
 - binding energy, 21–22, 27–28
 - bremsstrahlung radiation, 35–36
 - capture decay, nuclear transformation, 596–597
 - cascade, 23
 - cathode, source of, 97
 - elastic interaction, 34
 - inelastic interaction, 34
 - interactions, 34–36
 - orbit, 21–22
 - shell, 21–22
 - transition, 23–24
 - Auger electron, 23
 - characteristic x-ray, 23
 - electron cascade, 23
 - fluorescent yield, 24
 - radiation from, 23
 - valence shell, 22
 - Electronic gain, electron optics, fluoroscopy, 234
 - Electronic noise, 67
 - Electronic structure
 - atom, 21–22
 - binding energy, 21–22

- Electronic switch/timer, exposure timing, 132–133
- Elements useful in magnetic resonance imaging, characteristics of, 376–379, 376*t*, 377*f*, 378*f*, 379*t*
- Elevational resolution, and image quality, ultrasound, 500, 500*f*
- Emission, x-ray, factors affecting, 135–137, 136*f*, 137*f*
- Emission imaging, 7, 145
- Emulsion
 - number, 189
 - orthochromatic, 163
 - panchromatic, 163
- Energy
 - conservation of, 868
 - kinetic, 867
 - mechanical, 3
 - potential, 868
- Energy fluence, 52
- Energy resolution, definition, 637
- scintillation camera, 682
- Enhancement artifact, ultrasound, 527*f*, 529
- Equivalent dose, 56–57
- Ethernet, local area network, 559–560, 559*t*
- Even echo rephrasing, magnetic resonance imaging, 409
- Excitation, 32*f*
 - particle, 31, 32*f*
- Excited state, 24
- Exposure
 - measurement of, 636
 - definition, 54–56
 - rating chart, 140–144
 - timing, x-ray, 132–135
- Exposure control, 755–788
- Extended processing, 183, 183*f*, 217–219, 218*f*
- F**
- F-number, lenses, fluoroscopy, 239
- Fan-beam collimator, 676
- Fan beam geometry, x-ray computed tomography, 329
- Far field, ultrasound beam, 493
- Fast neutron, nuclear reactor produced radionuclide, 608
- Fast spin-echo acquisition, magnetic resonance imaging, 432–433
- Ferromagnetic materials, 881
 - defined, 374
- Fetus
 - radiation effects, in utero, 853–862
 - dose limits, 791*t*
 - dose from radiopharmaceuticals, 911
 - calculation of conceptus dose, 802–803
- FID. *See* Free induction decay
- Field of view
 - spatial resolution, x-ray computed tomography, 369
- Field of view, image intensifier, fluoroscopy, 236–237, 237*f*
- Field of view, nuclear medicine collimators, 695
- Filament circuit, x-ray, 126
- Fill factor, flat panel detectors, 302, 303*f*
- Film
 - characteristics of, 157–163, 157*f*
 - processing mammography, 212–219
- Film digitizers, image formats, 568–571, 569*f*, 570*f*
- Film exposure
 - development, 177
 - film emulsion, 175–176, 176*f*
 - latent image, 176–177
- Film processing, 175–189
 - Film-screen radiography, 145–173
- Film sensitometry, film processing, mammography, 215–217, 216*f*, 217*f*
- Filtered backprojection reconstruction, x-ray computed tomography, 330, 728
- Filtration, x-ray, 114, 115*t*
- Firewall, network security, 565
- Fission, 28–29, 607
- Fission-produced radionuclides, 609–611
- Fixer. *See* Film processor
- Flat field image, digital image processing, 309
- Flat panel display, 88–89, 577
- Flat panel detectors, 300–304
 - direct detection flat panel systems, 303–304, 304*f*
 - indirect detection flat panel systems, 300–303, 301*f*, 303*f*
- Flip angle, magnetic resonance imaging, 383–385
- Fluorine-18, decay scheme, 601–602, 602*f*
- Flow-related enhancement, magnetic resonance imaging, 408
- Flow turbulence, magnetic resonance imaging, 408
- Fluence, 52
- Fluorescence, scintillation detector, 637
 - light conversion of, into electrical signal, 639–641
- Fluoroscopy, 5
 - automatic brightness control, 246–248
 - brightness gain, image intensifier, 236
 - contrast ratio, image intensifier, 237
 - conversion factor, image intensifier, 236
 - detected quantum efficiency, image intensification, 238, 238*f*
 - equipment, 232–242
 - magnification mode, image intensification, 236–237
 - modes of operation, 244–246
 - overview, 233–234
 - peripheral equipment, 242–244

- pincushion distortion, artifact, image intensifier, fluoroscopy, 234–235
- radiation dose, 251–254
- S distortion, artifact, image intensifier, 238
- spatial resolution, limitation of, image intensification, 248
- suites, 249–251
- video camera, 240–242, 240*f*
- Flux, 52
- Focal spot
 - effect on spatial resolution, x-ray computed tomography, 368
 - mammography, 196–197, 197*f*, 197*t*
 - size, anode, 108–111, 108*f*, 111*f*
- Focusing cup, cathode, x-ray tubes, 102–103
- Fourier transform
 - and spatial resolution, 272–273, 272*f*
- FOV. *See* Field of view
- Frame averaging, fluoroscopy, 245–246, 245*f*
- Fraunhofer region, ultrasound beam, 493
- Free induction decay (FID) signal, magnetic resonance imaging, 385–387, 385*f*, 386*f*
- Frequency
 - domain and image quality, 269–270, 269*f*
 - electromagnetic wave, 17
 - encode gradient, magnetic resonance imaging, signal localization, 422–423
 - sound, 471–475
- Fresnel region, ultrasound beam, 490–493
- Fusion, 28–29
- G**
- Gamma camera. *See* Scintillation camera
- Gamma ray, 17, 27
- Gas-filled radiation detector, 632–636, 632*f*, 633
- Gastrointestinal radiation syndrome, 835
- Gated, frame mode acquisition, image acquisition, 696–697
- Gauss, 876
- Gaussian distribution, image quality, 274, 665
- Geiger-Mueller radiation counter, 635–636, 635*f*
- Generator
 - power, rating chart, 137–138, 138*f*
 - x-ray, 116–124
- Genetic effect, radiation, 851–853, 852*t*
- Geometric efficiency, radiation detection, 631
- Geometric orientation, magnetic resonance imaging, 379–381, 380*f*, 381*f*
- Geometric tomography, 317–320, 318*f*, 319*f*, 703
 - tomographic angle, 317
- Gibbs phenomenon, magnetic resonance imaging, 455, 456*f*
- Gonadal shielding, radiation protection, 775
- Gradient field artifacts, magnetic resonance imaging, 448, 449*f*
- Gradient moment nulling, 437, 437*f*
- Gradient recalled echo, 434–435, 434*f*
- Gradient sequencing, magnetic resonance imaging, 425
- Grating lobe artifact, ultrasound, 530
- Grating lobes, ultrasound beam, 495–497
- Gray, unit of absorbed dose, 53–54
- Gray-scale cathode ray tube monitor, 87–88, 87*f*, 89*f*
- Gray scale value, image quality, 260
- Grid frequency, 171
- Grid ratio, antiscatter grid, 169
- Grounded anodes, 195
- H**
- “H & D curve.” *See* Characteristic curve; Hurter and Driffield curve
- Half value layer (HVL), 48–52, 50*t*
 - mammography, 200–203, 202*f*, 203*t*
- Half-life
 - physical, 590–591, 590*f*, 591*f*
 - biological, 807
 - effective, 807
- Hard copy
 - device, for computer display, 93
 - display vs soft copy display, 308
 - laser multiformat camera, 581–582, 582*f*
 - multiformat video camera, 581, 581*f*
- Harmonized image, 314
- Heat loading, rating chart, x-ray tube, 139–140
- Heel effect, anode, 112, 112*f*
- Helical computed tomography, 337–338, 350–351, 350*f*, 351*f*
- Hematopoietic radiation syndrome, 833–835, 834*f*
- Hereditary effects of radiation exposure, 851–853
- Hertz, unit, 18
- High-velocity signal loss (HVSL), magnetic resonance imaging, 408
- Homogeneity coefficient, 51
- Hospital information system (HIS), 572
- Hounsfield unit, x-ray computed tomography, 356–357
- Housing cooling chart, 143–144, 144*f*
- Hurter and Driffield curve, 159–163, 159*f*, 162*f*
- Huygen's principle, ultrasound beam, 490
- HVL. *See* Half value layer
- I**
- Illuminance, film viewing conditions, mammography, 219
- Image acquisition, magnetic resonance imaging, 442
- Image co-registration, computer processing, 86, 735
- Image format
 - computer, 82–84, 83*t*
 - digital images, 568
- Image intensifier, fluoroscopy, 232–233

- Image magnification, spatial resolution, 266–268, 266*f*, 267*f*, 680
 - Image plates, computed radiography, 294
 - Image quality
 - and aliasing, 283–287, 284*f*, 285*f*
 - contrast, 255–263, 256*f*
 - contrast-detail curves, 287–288, 287*f*
 - contrast resolution, 248–249
 - detector contrast, 259–260, 260*f*
 - digital image contrast, 261–262
 - displayed contrast, 262–263, 262*f*
 - Fourier transform, 272–273, 272*f*
 - frequency domain, 269–270, 269*f*
 - limiting spatial resolution, 248
 - modulation transfer function, 270–271, 270*f*, 271*f*, 272*f*
 - noise, 273–283, 273*f*, 275*f*, 276*f*, 277*f*
 - frequency, 281–283, 282*f*
 - quantum noise, 278–279, 278*t*
 - radiographic contrast, 261, 261*f*
 - receiver operating characteristic (ROC) curves, 288–291, 288*f*, 291*f*
 - sampling, 283–287, 284*f*, 285*f*
 - secondary quantum sink, 280
 - spatial resolution, 248, 248*f*, 249*t*, 263–273
 - convolution, 265, 312–314, 312*f*, 313*f*
 - deconvolution, 314
 - delta function, 313
 - edge spread function, 269
 - kernel, 312
 - point spread function, 263–265, 264*f*, 265*f*
 - structured noise, 281
 - subject contrast, 256–259, 257*f*, 258*f*, 259*f*
 - temporal resolution, 249
 - Image reconstruction,
 - computed tomography, 351–355
 - filter, contrast resolution, x-ray computed tomography, 368, 369
 - magnetic resonance imaging, 426–437
 - SPECT and PET, 707–709, 727
 - Imaging
 - functional, 7
 - projection, 5
 - transmission, 5
 - Imparted energy, 56
 - Implementation, digital radiography, 307
 - In utero* radiation effects, 853–860
 - Inherent infiltration, x-ray, 114
 - Intensification factor, 150*f*, 153
 - Intensifying screens, 148
 - Intensity, sound, 475, 476*t*
 - Internal conversion electron, 27
 - Internal noise, interpretation, image quality, 290
 - International Commission on Radiological Protection (ICRP)
 - dosimetric quantity, 56, 57, 58, 58*t*
 - International System- of Units, 866, 866*t*
 - conventional unit equivalent, 57–58, 59*t*
 - radiation quantities, 57–58, 59*t*
 - Internet, 562
 - Intrinsic efficiency, radiation detection, 631
 - Intrinsic resolution, performance, scintillation camera, 679
 - Inversion recovery, magnetic resonance imaging, 399–401, 400*f*, 433
 - Inverter x-ray generator, 129–131, 130*f*
 - Iodine-125, spectrum of, 652–653, 652*f*
 - Ion pair, 31
 - Ionization, 32*f*
 - chamber, gas-filled radiation detector, 632–633, 634–634*f*
 - Ionizing radiation, 19
 - Isobaric transitions, beta-minus decay, 594
 - Isobars, 25*t*
 - Isocenter, computed tomography, 342
 - Isomeric state, 24
 - Isomeric transition, 27, 597
 - Isomers, 25*t*
 - Isotones, 25*t*
 - Isotopes, 25*t*
 - Iterative reconstruction
 - SPECT and PET, 708–709, 727
 - x-ray computed tomography, 351
- J**
- Joint Commission on the Accreditation of Health Organizations (JCAHO), 186
 - Joule
 - definition, 867
 - rating chart, x-ray tube, 139–140
- K**
- K shell, electron, 21–22
 - K-space data acquisition, magnetic resonance imaging, 426–437
 - K-space errors, magnetic resonance imaging, 450, 451*f*
 - Kerma, 52–53
 - Kernel, and spatial resolution, 312
 - Kinetic energy, 867
- L**
- Laboratory frame, magnetic resonance imaging, 379
 - LAN. *See* Local area network
 - Large flip angle, gradient recalled echo, magnetic resonance imaging, 406
 - Larmor equation, elements, nuclear magnetic characteristics, 377
 - Laser cameras
 - film processing, 184, 185*f*
 - Laser multiformat camera, 581–582, 582*f*
 - Last-frame-hold, fluoroscopy, 246
 - Latent image, film, 158, 177
 - Lateral resolution, and image quality, ultrasound, 498–500, 499*f*

Latitude, 163
 LCD. *See* Liquid crystal display
 Leakage radiation, 113
 LET. *See* Linear energy transfer
 Leukemia, radiation induced carcinogenesis, 849
 Line focus principle, 108
 Line spread function (LSF), 268, 679, 727
 Linear attenuation coefficient, 45–47
 Linear energy transfer (LET), 34
 Liquid crystal display (LCD), 88
 List-mode acquisition, scintillation camera, 698
 Local area network (LAN), 555, 558–561, 558*f*
 Localization of signal, magnetic resonance
 imaging, 415–425
 Longitudinal magnetization, magnetic resonance
 imaging, 380
 Look-up table, displayed contrast, image quality,
 90, 262
 Luminance
 video monitors, 577
 film viewing, mammography, 219
M
 Magnetic field, 874–876, 875*f*, 876*f*
 Magnetic field gradient, magnetic resonance
 imaging, signal localization, 415–418
 Magnetic field inhomogeneities, magnetic
 resonance imaging artifact, 391–393
 Magnetic field strength (magnetic flux density),
 magnetic resonance imaging, 374, 442
 Magnetic force, 874–876, 875*f*, 876*f*
 Magnetic moment, defined, magnetic resonance
 imaging, 375
 Magnetic resonance imaging, 9–12, 11*f*, 415–467
 acquisition time, 430
 ancillary equipment, 460–461, 460*f*, 461*f*
 angiography, 442–446, 443*f*
 artifacts, 447–457
 biological effects, 465–467, 466*t*
 blood oxygen level-dependent (BOLD), 409–410
 chemical shift artifacts, 453–455, 453*f*
 coil quality factor, 441–442
 contrast weighting parameters, spin-echo pulse
 sequence, 394–395, 394*f*
 cross-excitation, 442
 data synthesis, 431–432, 432*f*
 detection of signal, 381–385, 382*f*, 383*f*, 384*f*
 echo planar image acquisition, 435–436, 435*f*
 elements, characteristics of, 376–379, 376*t*,
 377*f*, 378*f*, 379*t*
 equilibrium, return to, 387–389, 387*f*, 388*f*,
 389*f*
 even-echo rephrasing, 409
 fast spin echo (FSE) acquisition, 432–433, 433*f*
 flip angle, 383–385
 flow-related enhancement, 408
 free induction decay signal, 385–387, 385*f*,
 386*f*

 frequency encode gradient, signal localization,
 422–423, 422*f*, 423*f*
 generation of signal, 381–390
 geometric orientation, 379–381, 380*f*, 381*f*
 gradient field artifacts, 448, 449*f*
 gradient moment nulling, 437, 437*f*
 gradient recalled echo, 403–407, 404*f*
 gradient recalled echo acquisition, 434–435,
 434*f*
 gradient sequencing, 425, 425*f*
 image acquisition, 442
 image characteristics, 439–442
 image reconstruction, 426–437
 inversion recovery, 399–401, 400*f*
 inversion recovery acquisition, 433, 434*f*
 “k-space” data acquisition, 426–437, 426*f*
 k-space errors, 450, 451*f*
 laboratory frame, 379
 large flip angle, gradient recalled echo, 406
 localization of signal, 415–425
 longitudinal magnetization, 380
 machine dependent artifacts, 447
 magnetic field
 exposure limit, 467
 noise limit, 467
 magnet, 458–460
 permanent, 460
 resistive, 458–459
 superconductive, 459, 459*f*
 magnetic field
 gradient, signal localization, 415–418, 416*f*,
 417*f*, 418*t*
 inhomogeneities, 391–393
 magnetic field strength, 442
 magnetic moment, defined, 375
 magnetization properties, 373–381, 374*f*
 magnetization transfer, 411
 magnetization transfer contrast, 443*f*, 445–446
 motion artifacts, 451–453, 451*f*, 452*f*
 multislice data acquisition, 431, 431*f*
 nucleus, characteristics of, 375, 375*t*
 pairing phenomenon, 375
 partial-volume artifacts, 457
 phase encode gradient, signal localization,
 424–425, 424*f*
 quality control, 462–465, 464*f*, 464*t*
 radiofrequency artifacts, 449–450
 radiofrequency coil artifacts, 449, 450*f*
 reconstruction algorithms, 442
 RF bandwidth, 440–441, 441*f*
 ringing artifact. *See* Gibbs phenomenon
 rotating frame, 379, 383
 safety, 465–467, 466*t*
 signal averages, 440
 signal-to-noise ratio, 440
 slice encode gradient, signal localization,
 418–421, 418*f*, 419*f*, 420*f*, 421*f*
 small flip angle, gradient recalled echo, 406

- Magnetic resonance imaging (*contd.*)
- spatial resolution, with contrast sensitivity, 439
 - spin-echo, 391–399
 - spin-echo pulse sequence, 401–402, 403*f*
 - spin-lattice relaxation, defined, 387
 - spin-spin interaction, 386
 - spiral k-space acquisition, 436–437, 436*f*
 - static magnetic field, biological effect, 466
 - susceptibility artifacts, 447–448
 - T1, T2 value, tissues, 389–390, 390*f*, 391*t*
 - three-dimensional Fourier transform imaging,
 - signal localization, 438–439, 438*f*
 - time of echo, spin-echo, 391–393, 392*f*, 393*f*
 - time-of-flight angiography, 443–445, 444*f*
 - time of repetition, spin-echo, 393, 394*f*
 - transverse magnetization, 380
 - two-dimensional data acquisition, 426–430,
 - 427*f*, 428*f*, 429*f*
 - two-dimensional multiplanar acquisition, 430,
 - 430*f*
 - varying magnetic field effects, 466
 - voxel volume, 440
 - wrap-around artifacts, 455–457, 457*f*
- Magnetization properties, magnetic resonance imaging, 373–381, 374*f*
- Magnetization transfer, magnetic resonance imaging, 411, 445–446
- Magnification mode, image intensification, fluoroscopy, 236–237
- Magnification in nuclear imaging, 675–676,
 - 680–681, 687, 687*f*
- Magnification technique, mammography,
 - 210–212, 211*f*
- Mammography, 5–6, 6*f*, 191–230
- anode, 194–195, 195*f*, 196*f*
 - automatic exposure control, x-ray generator,
 - 204–206, 205*f*
 - average glandular dose, radiation dosimetry,
 - 222–223
 - back-up timer, 206
 - Bucky factor, 210
 - cathode, 194
 - collimation, 203–204
 - compression of breast, 207, 208*f*
 - film processing, 212–219
 - extended cycle processing, 217–219,
 - 218*f*
 - film processor types, 212–219
 - film viewing conditions, 219
 - focal spot, 196–197, 197*f*, 197*t*
 - half value layer, 200–203, 202*f*, 203*f*
 - magnification technique, 210–212, 211*f*
 - mid-breast dose, radiation dosimetry, 222
 - molybdenum, 194
 - pinhole camera, 196
 - quality assurance, 225–229
 - radiation dosimetry, 222–224
 - reciprocity law failure, 206
 - regulatory requirements, 224–229
 - rhodium, 194, 198
 - scattered radiation, antiscatter grid, 207–210,
 - 209*f*
 - single screen, single film emulsion, 212–213
 - slit camera, 196
 - space charge effect, 194
 - technique chart, 206–207, 206*f*
 - tube port, filtration, 197–200, 198*f*, 199*f*,
 - 201*f*, 202*f*
 - x-ray generator, phototimer system, 204–207
 - x-ray tube design for, 113, 194–204
- Mammography Quality Standards Act (MQSA), 191
- Mass attenuation coefficient, 45*f*, 47–48, 47*f*
- Mass defect, 27–28
- Mass energy absorption coefficient, 54
- Mass energy transfer coefficient, 53
- Mass number, 24
- Matching layer, ultrasound, 487–488
- Maximum intensity projection (MIP), x-ray
 - computed tomography, 362
- Mean free path (MFP), 50
- Mean glandular dose, breast. *See* Average glandular dose
- Medical internal radiation dosimetry (MIRD)
 - method, of radiopharmaceutical dosimetry,
 - 805–812, 806*f*
 - accuracy of dose calculations, 811–812
 - cumulated activity, 807–808, 807*f*, 808*f*
 - dose calculation, example of, 810
 - S factor, 808–810, 809*t*, 810*t*
- Memory, computer, 70–72, 71*f*
- Metastable state, 24
 - nuclear transformation, 597
- Minification gain, output phosphor, 236
- MIPS. *See* Millions of instructions per minute
- MIRD. *See* Medical internal radiation dosimetry
- Mirror image artifact, ultrasound, 530
- Mo-99/Tc-99m radionuclide generator, 612–614,
 - 613*f*
 - quality control, 616–617
 - molybdenum breakthrough, 616–617
- Modem, 73, 77, 81, 563
- Modulation transfer function (MTF)
 - and image quality, 270–271, 270*f*, 271*f*, 272*f*
- Molybdenum-99, decay scheme, 598–600, 600*f*
- Molybdenum breakthrough, Mo-99/Tc-99m
 - radionuclide generator, 616–617
- Molybdenum concentration, measured, dose
 - calibrator, 660–661
- Momentum, defined, 868
- Motion artifacts, magnetic resonance imaging,
 - 451–453, 451*f*, 452*f*
- MQSA. *See* Mammography Quality Standards Act
- MRI. *See* Magnetic resonance imaging
- MSAD. *See* Multiple scan average dose
- MTF. *See* Modulation transfer function

Multi-path reflection artifact, ultrasound, 530
 Multichannel analyzers (MCAs) system, pulse height spectroscopy, 647–648, 647*f*, 648*f*
 Multienergy spatial resolution, 681
 Multiformat video camera, 581, 581*f*
 Multiple detector arrays. x-ray computed tomography, 341–342, 341*f*
 Multiple exposure, rating chart, 142
 Multiple scan average dose (MSAD), dose measurement, x-ray computed tomography, 363
 Multiplication factor, nuclear reactor, 608
 Multislice data acquisition, magnetic resonance imaging, 431

N

Narrow-beam geometry, 48–49, 48*f*
 National Council on Radiation Protection and Measurement (NCRP), 789
 National Electrical Manufacturers Association (NEMA), specifications, anger scintillation camera, 571, 690–691
 Nationwide Evaluation of X-Ray Trends (NEXT), 797
 Near field, ultrasound beam, 490–493, 491*f*, 492*f*
 Neurovascular radiation syndrome, 835–836
 Noise, and image quality, 273–283, 273*f*, 275*f*, 276*f*, 277*f*
 Noise frequency, and image quality, 156–157, 281–283, 282*f*
 NRC. *See* Nuclear Regulatory Commission
 Nuclear fission, 28–29
 nuclear reactor produced radionuclide, 606–607, 607*f*
 Nuclear forces, atom, 24
 Nuclear reactors, 608–609, 608*f*, 610*f*
 Nuclear reactor produced radionuclide, 606–611, 607*f*
 Nuclear Regulatory Commission (NRC), 783, 788–791
 Nuclear transformation, 593–602
 alpha decay, 593
 beta minus decay, 594–595, 594*f*
 beta-plus decay, 595–596, 596*f*
 decay schemes, 598–602, 598*f*
 fluorine-18, 601–602, 602*f*
 molybdenum-99, 598–600, 600*f*
 phosphorus-32, 598, 599*f*
 radon-220, 598, 599*f*
 technetium-99m, 600–601, 601*f*
 electron capture decay, 596–597
 isomeric transition, 597
 metastable state, 597
 Nucleon, 24
 Nucleus. *See* Atomic nucleus characteristics of, magnetic resonance imaging, 375, 375*t*

Nuclide classification, 25

Number systems

 binary form, 62, 62*t*
 conversion, decimal form, binary form, 62–63, 63*t*

 Nyquist frequency, 68, 302

O

Off-focus radiation, anode, 112

Optical coupling

 fluoroscopy, 238–239, 239*f*

Optical density, light transmission relationship, 158, 158*t*

Organ response to radiation, 827–837

 ocular effects, 831

 radiosensitivity, 828

 regeneration/repair, 827–828, 828*f*

 reproductive organs, 831

 skin, 829–830, 830*t*

Organogenesis, radiation effects on the fetus, 854–856, 855*f*

P

PACS. *See* Picture archiving and communication system

Pair production, 44, 44*f*, 45*f*

Pairing phenomenon, nucleus, magnetic characteristics, 375

Par speed system, 161

Parallel beam geometry, x-ray computed tomography, 329

Parallel-hole collimator, 674

Parallel processing, 73

Paralyzable model, of pulse mode, for radiation detection, 629, 629*f*

Paramagnetic materials, 374, 880–881

Parent nuclide, 27

Partial volume effect

 magnetic resonance imaging artifact, 457

 ultrasound, 647

 x-ray computed tomography, 371–372, 372*f*

Particulate radiation, 20–21

Path length, particle, 33

PEG. *See* Phase encode gradient

Perfusion contrast, magnetic resonance imaging, 409–411

Permanent magnet, magnetic resonance imaging, 460

Personnel dosimetry, 747–753

 film badge, 748–750, 749*f*

 pocket dosimeter, 751–753, 752*f*, 753*t*

 thermoluminescent dosimeter (TLD), 750, 751*f*

PET. *See* Positron emission tomography

Phase contrast angiography, magnetic resonance imaging, 445–446*f*

Phase encode gradient (PEG), magnetic resonance imaging, signal localization, 424–425

- Phased array, ultrasound, 490
 Phosphorus-32, decay scheme, 598, 599*f*
 Photodiode
 computed tomography, 340
 fluoroscopy, 247
 scintillation detector, 640–641
 Photoelectron, 40
 Photoelectric absorption, 40–43, 41*f*, 43*f*, 45*f*
 Photomultiplier tube (PMT)
 exposure timing, 133
 positron emission tomography (PET),
 721–724, 722*f*
 scintillation camera, 671–672
 scintillation detector, 639–640, 640*f*
 Photon, 17
 Photopeak, 650–653, 658
 Photopeak efficiency, scintillation camera, 682
 Photostimulable phosphor, 641
 Phototimer system
 exposure timing, 133–134, 134*f*
 mammography, 204–207
 Picture archiving and communications systems
 (PACS), 565–584
 Piezoelectric crystal, ultrasound, 483–484, 485*f*,
 486*f*
 Pincushion image distortion, in fluoroscopy,
 234–235
 Pinhole camera, focal spot size 110, 196
 Pinhole collimator, 675
 PMT. *See* Photomultiplier tube
 Point spread function (PSF), 263–265
 Polyenergetic x-ray beam, 50
 Portable ionization chamber, radiation protection,
 634*f*, 754–755, 755*f*
 Positive predictive value 290
 Positron emission tomography (PET), 9, 10*f*, 703,
 719–735, 719*f*
 Potential energy, 868
 Power Doppler, 543–544
 Pressure, sound, 475, 476*t*
 Primary barrier, radiation protection, 763
 Primordial radionuclides, 740–741
 Probability density functions (PDFs), 275,
 662–663
 Processing
 daylight, 184
 dry, 184–185
 extended, 183, 183*f*
 rapid, 183
 Processor quality assurance, 186–189
 Progression, cancer development stages, 839
 Projection image
 tomographic image vs, 5, 6 22, 327–328, 703
 Projection radiology 154–162
 absorption efficiency, 154–155
 average gradient, 160, 160*f*
 basic geometric principles, 146–148, 146*f*,
 147*f*
 characteristic curve of film, 159–163, 159*f*,
 162*f*
 conversion efficiency, 153–154
 densitometer, 158
 latitude, 162–163
 orthochromatic emulsion, 163
 overall efficiency, 155–157
 panchromatic emulsion, 163
 par speed system, 161
 projection image, defined, 145–146*f*
 quantum detection efficiency (QDE), 151
 reciprocity law, 163–164
 scattered radiation, 166–173, 167*f*, 168*f*
 screen-film system, 163–164
 transmission, 145
 transmittance, 158
 Proportional counter, gas-filled radiation detector,
 635
 Proportional region, gas-filled radiation detector,
 633
 PSF. *See* Point spread function
 Pulse echo operation, ultrasound, 502–504, 503*f*,
 504*t*
 Pulse height analyzers (PHAs), 645
 multichannel analyzer, 645
 single channel analyzer, 645–647, 645*f*, 646*f*
 Pulse height spectroscopy, 644–654
 cesium-137, spectrum of, 650–651, 650*f*
 iodine-125, spectrum of, 652–653, 652*f*
 technetium-99m, spectrum of, 651–652, 651*f*
 Pulse mode, radiation detection, 628–630
 non-paralyzable model, 629, 629*f*
 paralyzable model, 629, 629*f*
 Pulse sequences, magnetic resonance imaging, 391
 Pulsed fluoroscopy, radiation protection, 778
 Push processing, film processing, 183
- Q**
 QDE. *See* Quantum detection efficiency
 Quality assurance, *See* quality control
 Quality control
 dose calibrator, 658
 mammography, 225–229
 magnetic resonance imaging, 462–465, 464*f*,
 464*t*
 scintillation, 691–692
 SPECT, 714
 thyroid probe, 658
 well counter, 658
 Quality Factor
 radiation, 57
 ultrasound, 487
 Quanta, 17
 Quantities, radiation. *See* Radiation quantities
 Quantization, data conversion, 68
 Quantum accounting diagram, 279*f*, 280
 Quantum detection efficiency (QDE)
 radiation detection, 630–631, 631*f*

- for screen film system, 151
- Quantum levels, electron, 21–22
- Quantum noise
 - and image quality, 278–279, 278*t*
- Quantum number, electron shell, 21–22

R

Rad

- definition, 54
- roentgen to rad conversion factor, 54, 55*f*

Radiation

- absorption, 17
- electromagnetic spectrum, 17, 18*f*
- electromagnetic wave characteristics, 17–18, 18*f*
- electromagnetic radiation, 17–19, 19*f*
- electron interaction, 34–35
 - bremstrahlung radiation, 35–36
 - elastic, 34
 - inelastic, 34
- electron transition, 23–24
- frequency, 17
- gamma ray, 17
- interaction with matter, 17–59
 - attenuation, electromagnetic radiation, 45–52
 - electromagnetic radiation interaction, 34–35
 - gamma ray interactions, 37–44
 - particle interactions, 31–37
- ionizing radiation, 19
- ionization, defined, 627
- leakage, 113
- neutron interactions, 36–37, 36*f*
- particle characteristics, 19
 - photon, 17
 - quanta, 17
- particle interaction
 - annihilation, 9, 596, 719
 - Bragg peak, 32–33
 - delta rays, 31, 32*f*
 - excitation, 31, 32*f*
 - ion pair, 31
 - ionization, 31, 32*f*
 - linear energy transfer, 34
 - radiative loss, 31
 - secondary ionization, 31
 - specific ionization, 31, 33*f*, 34
- particulate radiation, 20–21
- photon, 17
- quanta, 17
- quantities and units, 52–56
 - absorbed dose, 53–54
 - effective dose, 57–58
 - equivalent dose, 56–57
 - exposure 54–56
 - fluence, 52
 - flux, 52
 - gray, 52–53
 - integral dose, 57

- kerma, 52–53
 - air, 52–53
- rad, 53–54
- radiation weighing factor, 56–57, 57*t*
- rem, 58
- roentgen, 53
- roentgen to rad conversion factor, 54, 55*f*
- sievert, 57
- radio frequency, 17
- scattering, 17
- types of, 17
- unchanged particle interaction, 36–37, 36*f*
- units. *See* Quantities, radiation
- weighting factor, 56–57
- wavelength, 17
- x-ray, 17

Radiation biology

- acute radiation syndrome, 831–837
- biological effects
 - classification of, 814
 - determinants of, 813–814, 814*t*
 - deterministic, 795
 - stochastic, 795
- breast cancer, 850–851
- cancer development, stages of, 839–840
- cellular radiobiology, 818–827
 - cell survival curves, 821–822, 822*f*
 - cellular radiosensitivity, 821
 - cellular targets, 818
 - DNA, 818–821, 819*f*, 820*f*
- epidemiological investigation, 840–841, 841*t*, 842*t*
- gastrointestinal radiation syndrome, 835
- genetic effect, radiation, 851–853, 852*t*
- hematopoietic radiation syndrome, 833–835, 834*f*
- hereditary effects of radiation exposure, 851–853
- leukemia, 849
- neurovascular radiation syndrome, 835–836
- organ response to radiation, 827–837
 - ocular effects, 831
 - radiosensitivity, 828
 - regeneration/repair, 827–828, 828*f*
 - reproductive organs, 831
 - skin, 829–830, 830*t*
- radiation induced carcinogenesis, 838–851
- risk estimation models, 841–848, 844*f*, 845*f*, 846*f*, 847*t*, 848*t*
- thyroid cancer, 849–850
- tissue/radiation interaction, 814–818, 816*f*, 817*f*
- in utero, radiation effects, 853–860. *See In utero* radiation effects

Radiation detection

- current mode, 628–630
- dead-time, 628
- detector system, 628
- detector types, 627–631
- dosimeter, 628

- Radiation detection (*contd.*)
 - gas-filled radiation detector, 632–636
 - Geiger-Mueller counter, 635–636
 - ionization chamber, 634
 - proportional counter, 635
 - efficiency, detection, 630–631, 631*f*
 - pulse height spectrum, 630, 630*f*
 - pulse mode, 628–630
 - non-paralyzable model, 629, 629*f*
 - paralyzable model, 629, 629*f*
 - scintillation detector. *See* Scintillation detector
 - semiconductor detector, defined, 628
 - spectrometers, defined, 628
 - spectroscopy, defined, 630, 630*f*
 - thermoluminescent dosimeter, 641
- Radiation dose. *See* Dose
- Radiation dosimetry. *See* Dose
- Radiation induced carcinogenesis. *See* Radiation biology
- Radiation protection
 - advisory bodies
 - International Commission on Radiological Protection, 56, 57, 58, 58*t*
 - National Council on Radiation Protection and Measurement, 789
 - equipment, 753–755
 - Geiger-Mueller survey instruments, 754
 - portable ionization chamber, 754–755, 755*f*
 - exposure control. *See* Exposure control
 - personnel dosimetry
 - film badge, 749–750, 749*f*
 - pocket dosimeter, 751–753, 752*f*, 753*t*
 - problems with, 753
 - thermoluminescent dosimeter (TLD), 750, 751*f*
 - regulatory agencies, 788–789
- Radiation weighing factor, 56–57, 57*t*
- Radiopharmaceuticals, 617–624
 - localization mechanisms, 620–623
 - physical characteristics of, 618*t*–619*t*
 - quality control, 617
 - safety, 620
- Radioactive decay, 25, 589
 - activity defined, 589, 590*t*
 - decay constant, defined, 589–590
 - fundamental decay equation, 591–593, 592*f*
 - physical half-life, 590–591, 590*t*, 591*t*
 - terminology, 589–593
- Radioactivity, 25–29
- Radiographic contrast, 261, 261*f*
- Radiography, 4–5
- Radiology information system (RIS), 572
- Radionuclide
 - modes of decay, 589
 - production, 603
- Radionuclide decay. *See* Radioactive decay
- Radionuclide generator
 - Mo-99/Tc-99m radionuclide generator, 612–614, 613*f*
 - quality control, 616–617
 - secular equilibrium, 615–616, 615*t*, 616*f*
 - transient equilibrium, 614–615, 614*f*
- Radionuclide production, 603–617
 - cyclotron-produced radionuclide, 603–606, 604*f*, 606*f*
 - nuclear reactor-produced radionuclide, nuclear fission, 606–611, 607*f*
- Radionuclide therapy 784–787
- Radiopharmaceuticals
 - dose calibration, 659–660
 - dose from, 708–711
 - dosimetry, medical internal radiation dosimetry (MIRD) method, 805–812, 806*f*
 - quality control, 623–624
 - regulation of, 624–626
 - U.S. Food and Drug Administration, regulation, 624
 - U.S. Nuclear Regulatory Commission, regulation, 625
- Radon-220
 - decay scheme, 598, 599*f*
- Random access memory (RAM), 71
- Random error, counting statistics, 661
- Range, particle, 33
- RAP meters. *See* Roentgen-area product meters
- Rare earth, screen, 149
- Rating chart
 - generator power, 137–138, 138*t*
 - multiple exposure, 142
 - single exposure, 140–142, 141*f*
 - x-ray tube, 137–138, 139*t*
 - anode cooling chart, 142–143, 142*f*
 - anode heat input chart, 142–143, 142*f*
 - heat loading, 139–140
 - housing cooling chart, 143–144, 144*f*
 - joule, 139–140
 - power rating, 137
- Rayleigh scattering, 37–38, 37*f*, 45*f*
- Read-only memory, (ROM) 71
- Receive focus, ultrasound beam, 494–495, 495*f*
- Receiver operating characteristic (ROC) curves, and image quality, 288–291, 288*f*, 291*f*
- Reciprocity, law, 163–164
 - failure, mammography, 206
- Reconstruction algorithm. *See* Image reconstruction
- Reconstruction filter. *See* Image reconstruction
- Rectifier circuit, x-ray, 124–126, 125*f*
- Recursive filtering, frame averaging, fluoroscopy, 246
- Reflection, ultrasound, 477–479, 478*f*, 479*t*
- Refraction
 - artifacts, 526, 527*f*
 - ultrasound, 479–480
- Relative biologic effectiveness (RBE), 817
- Relative risk (RR), radiation biology, 845
- Rem, 58

- Resistive magnet, magnetic resonance imaging, 458–459
 - Resonance transducers, 484–486
 - Reverberation, ultrasound, 528*f*; 529–530
 - Reverse bias, photodiodes, 640, 644
 - RF bandwidth, magnetic resonance imaging, 440–441, 441*f*
 - Rhodium, mammography, 194, 198
 - Ring artifact, magnetic resonance imaging. *See* Gibbs phenomenon
 - Road mapping, fluoroscopy, 246
 - ROC. *See* Receiver operating characteristic
 - Roentgen
 - definition, 54–55
 - to rad conversion factor, 54, 55*f*
 - Roentgen-area product meters, 253–254
 - ROM. *See* Read-only memory
 - Rose's criterion, contrast resolution, image quality, 280
 - Rotating anode, 106, 107*f*
 - Rotating frame, magnetic resonance imaging, 379, 383
 - Rotor/stator, x-ray tubes, 102, 106
 - Runback, processor artifacts, 183
 - S**
 - S distortion, artifact, image intensifier, fluoroscopy, 238
 - Sampling
 - analog to digital converters, 68–69
 - image quality and, 283–287, 284*f*; 285*f*
 - Sampling aperture, computed tomography, 351
 - Sampling pitch
 - computed tomography, 351
 - image quality, 283
 - Scatter degradation factor, mammography, 207
 - Scattered radiation
 - antiscatter grid, 168–173, 168*f*
 - mammography, 207–210, 209*f*
 - in projection radiology system, 166–173, 167*f*, 168*f*
 - in nuclear imaging 676, 691
 - reduction of, 168
 - scattered coincidences, 720, 721, 724
 - Scattering
 - radiation, 17, 34
 - ultrasound, 480–481, 480*f*
 - Scintillation camera, 669–702
 - collimator, 674–676, 675*f*
 - efficiency, 682
 - resolution, 679
 - design, 671–678
 - collimator resolution, and collimator efficiency, 684–686, 685*f*; 686*t*
 - intrinsic spatial resolution, and intrinsic efficiency, 683–684, 684*f*
 - spatial linearity, and uniformity, 688–690, 688*f*; 689*f*; 690*f*
 - system spatial resolution, and efficiency, 686–688, 687*f*
 - detector, 671–674, 672*f*; 673*f*; 674*f*
 - electronics, 671–674, 672*f*; 673*f*; 674*f*
 - energy resolution, 682
 - image formation principles, 676–678, 677*f*; 678*t*
 - intrinsic efficiency, 682
 - intrinsic resolution, 679
 - multienergy spatial registration, 681
 - multiple window spatial registration, 681
 - National Electrical Manufacturers Association (NEMA), specifications, 690–691
 - operation/quality control, 691–694, 692*f*, 693*f*
 - performance, 678–691
 - measures, 678–683
 - photopeak efficiency, 682
 - spatial linearity, 681
 - spatial resolution, 679
 - system efficiency, 681–682
 - uniformity, 678–679
- Scintillation detector, 636–641
 - afterglow, defined, 637
 - excited electron, 641
 - thermoluminescent dosimeters, 641
- fluorescence, 637
- inorganic crystalline scintillator, 638–639, 639*t*
- luminescence, 637
- phosphorescence, defined, 637
- photodiode, 640–641
- photomultiplier tube, 639–640, 640*f*
- photostimulable phosphors, 293, 641
- Screen-film combination, screen-film system, 163
- Screen-film system
 - and projection radiology, 163–164, 163*f*
- Screens
 - characteristics of, 149–157
 - composition and construction, 149–150
 - intensifying, 150–153, 151*f*; 152*f*
- Secondary ionization, 31
- Secondary quantum sink, and image quality, 280
- Sector scanning, ultrasound imaging, 13
- Secular equilibrium, 615–616, 615*t*, 616*f*
- Security
 - computer, 80–81
 - network, 564–565
- SEG. *See* Slice encode gradient
- Segmentation, three-dimensional image display, x-ray computed tomography, 360
- Semiconductor detector, radiation detection, 641–644, 642*f*; 643*f*
- Sensitivity speck, film processing, 176
- Sensitometer
 - defined, 186, 186*f*
- Septa, scintillation camera collimator, 672
- Shadowing artifact, ultrasound, 527*f*; 529
- Shielding, from radiation, 758–773
 - diagnostic radiology, 758–771
 - nuclear medicine, 771–773

- Shoe marks, processor artifacts, 182
- Side lobe, ultrasound beam, 495–497
- Side lobe artifact, ultrasound, 530
- Sievert, 57
- Signal averages, magnetic resonance imaging, 440
- Signal averaging, image quality, 284
- Signal-to-noise ratio
 - magnetic resonance imaging, 440
- Signal-to-noise ratio data conversion, 69
- Similar triangles, basic geometric principles, 146
- Simple backprojection, x-ray computed tomography, 352–353
- Single channel analyzers (SCAs), pulse height spectroscopy, 645–647, 645*f*, 646*f*
- Single exposure, rating chart, 140–142, 141*f*
- Single film emulsion, single screen, mammography, 212–213
- Single phase x-ray generator, 126–127, 126*f*, 127*f*
- Single photon emission computed tomography (SPECT), 7–9, 8*f*, 704–719
 - attenuation correction in, 709–711, 710*f*
 - collimator, 711, 711*f*
 - multihead cameras, 711–712
 - phantoms, 717–719, 717*f*, 718*t*
 - positron emission tomography, compared, 734, 734*t*
 - quality control, 714
 - transverse image reconstruction, 707–709, 708*f*, 709*f*
 - uniformity, 715–716, 716*f*
- Sinogram, tomographic reconstruction, 346–348, 718*t*
- Skin, response to radiation, 829–830, 830*t*
- Slap-line, processor artifacts, 182
- Slice encode gradient (SEG), magnetic resonance imaging, signal localization, 418–421
- Slice sensitivity profile
 - spatial resolution, x-ray computed tomography, 343, 368
- Slice thickness
 - contrast resolution, x-ray computed tomography, 369
 - single detector array scanners, computed tomography, 342–343
 - spatial resolution, x-ray computed tomography, 368
 - ultrasound, 531
- Slip-ring technology, helical scanner, x-ray computed tomography, 337
- Slit camera
 - anode, 110
 - mammography, 196
- Small flip angle, magnetic resonance imaging, 406
- Smoothing, spatial filtering, computer image processing. *See* Spatial filtering
- SNR. *See* Signal-to-noise ratio
- Sodium iodide scintillator, 627, 671
- Sodium iodide thyroid probe/well counter, 654–658, 655*f*
- Sodium iodide well counter, 656–657
- Soft copy display, vs hard copy display, 308
- Software, 78–81
 - applications program, 80
 - computer languages, 79
 - high-level language, 79, 80*f*
 - machine language, 79
 - operating system, 80
- Solid-state detector, x-ray computed tomography, 340–341, 340*f*
- Sound
 - attenuation 481–483, 481*f*, 482*f*
 - characteristics of, 470–475
 - decibel, 475, 476*t*
 - frequency, 471–475
 - intensity, 475, 476*t*
 - longitudinal wave, 470
 - mechanical energy, propagation, 470
 - pressure, 475, 476*t*
 - reflection 477–479, 478*f*, 479*t*
 - refraction 479–480
 - scattering 480–481, 480*f*
 - speed of, 471–475, 472*t*
 - wavelength, 471–475, 473*f*
- Space charge effect
 - cathode, 104
 - mammography, 194
- Space charge limited, cathode, x-ray tubes, 104
- Spatial filtering, 85, 312, 699
- Spatial resolution, 14–15
 - with contrast sensitivity, magnetic resonance imaging, 439
 - convolution, 265
 - deconvolution, 314
 - delta function, 313
 - edge spread function, 269
 - Fourier transform, 272–273, 272*f*
 - image intensification, fluoroscopy, 248, 248*f*, 249*t*
 - and image quality, 263–273
 - kernel, 312
 - limiting, 15*t*
 - line spread function, 268, 679, 727
 - MTF, 271, 273
 - point spread function, 263–265, 264*f*, 265*f*
 - ultrasound beam, 497–500
 - vs contrast resolution, digital imaging, 315–316, 315*f*
 - x-ray computed tomography, 368–369
- Specific ionization, 31, 33*f*
- SPECT. *See* Single photon emission computed tomography
- Spectroscopy
 - defined, 630, 630*f*
 - radiation detection, 630, 630*f*
- Speed, of sound, 471–475, 472*t*

Speed artifact, ultrasound, 528*f*, 530
 Spin echo
 magnetic resonance imaging, 10
 Spin-echo, magnetic resonance imaging, 391–399
 Spin-lattice relaxation, defined, magnetic resonance imaging, 387
 Spin-spin interaction, magnetic resonance imaging, 386
 Spiral k-space acquisition, 436–437, 436*f*
 “Spoiled” gradient echo technique, magnetic resonance imaging, 407, 408*f*
 Spot-film devices, 234
 Standard deviation, image quality, 274
 Star pattern, anode, 110
 Static, frame mode acquisition, image acquisition, 696–697
 Static magnetic field, biologic effect, magnetic resonance imaging, 466
 Stationary anode, 106
 Stereotactic breast biopsy, 219–221, 220*f*
 Storage of images, 573–575, 574*f*
 data compression, 575–576, 575*t*
 image management, 575
 Storage phosphors, computed radiography, 294
 Structure, atom. *See* Atom structure
 Structured noise, 281
 Subject contrast, and image quality, 256–259, 257*f*, 258*f*, 259*f*
 Sum peak, radiation spectrum 653
 Superconductive magnet, magnetic resonance imaging, 459, 459*f*
 Surface rendering, three-dimension image display, x-ray computed tomography, 361
 Susceptibility artifacts, magnetic resonance imaging, 447–448
 System efficiency, scintillation camera, 681
 Systematic error, counting statistics, 661

T

T1 weighting, magnetic resonance imaging, 395–396
 T2 weighting, magnetic resonance imaging, 397–398
 Target theory, cellular radiobiology, 818
 TDI. *See* Time delay and integration
 Technetium-99m
 decay scheme, 600–601, 601*f*
 spectrum of, 651–652, 651*f*
 Technique chart, mammography, 206–207, 206*t*
 Teleradiology, 568
 quality control of, 584–585, 584*f*
 Temporal resolution, fluoroscopy, 249
 Temporal subtraction, 321–323, 322*f*
 digital subtraction angiography, 321–323, 322*f*
 TFT. *See* Thin film transistor
 Thermionic emission cathode, 103
 Thermoluminescent dosimeter (TLD)
 radiation detection, 641
 scintillation detector, 641
 Thin film transistor (TFT)
 displays, 90
 flat panel radiographic detector, 301
 Three-dimensional Fourier transform imaging, magnetic resonance imaging, signal localization, 438–439, 438*f*
 Three-phase x-ray generator, 127–129, 128*f*
 Thermoluminescent dosimeter (TLD), 641, 750, 751*f*
 Thyroid cancer, radiation induced carcinogenesis, 849–850
 Thyroid probe-well counter, sodium iodide, 654–658, 655*f*
 Thyroid shields, personnel protection in diagnostic radiology, 771
 Thyroid uptake measurements, 655
 Time-activity curve (TAC), scintillation camera, 699
 Time delay and integration (TDI), digital mammography, 307
 Time-of-flight angiography, magnetic resonance imaging, 443, 445
 Time of inversion (TI), magnetic resonance imaging, 399
 Time of repetition (TR), magnetic resonance imaging, 393
 TLD. *See* Thermoluminescent dosimeter
 Tomographic angle, geometric tomography, 317
 Tomographic emission images, 9
 Tomographic image, 6
 magnetic resonance imaging, 10
 projection image vs, 6
 Tomographic reconstruction. *See* Image reconstruction
 Transducer, ultrasound, 483–490, 484*f*
 Transformers, x-ray, 116–119, 117*f*, 118*f*
 Transient equilibrium, 614–615, 614*f*
 Transmission control protocol/internet protocol (TCP/IP), 561–563
 Transmission imaging, 5, 145
 mammography, 5
 Transmit focus, ultrasound beam, 494, 494*f*
 Transmit/receive switch, image data acquisition, ultrasound, 502
 Transmittance, optical density, film, 158
 Transverse magnetization, magnetic resonance imaging, 380
 Triode, x-ray, 121
 Tube, x-ray. *See* X-ray, tube
 Tube output, 203, 204*f*
 Tube port, filtration, mammography, 197–200, 198*f*, 199*f*, 201*f*, 202*f*
 Tungsten atom
 de-excitation of, 23*f*
 Two dimensional data acquisition, magnetic resonance imaging, 426–430

Two dimensional multiplanar acquisition,
magnetic resonance imaging, 430
"Two-pulse" waveform, x-ray generator circuit
designs, 126

U

Ultrasound, 469–553, 470*f*. *See* Doppler
ultrasound
A-mode, transducer, 508
acoustic impedance, 14, 477, 477*t*
acoustic power, pulsed ultrasound intensity,
549–551, 549*f*, 550*t*
ambiguity artifact, 504
analog to digital conversion, 504–505, 505*f*
array transducer, 489–490, 489*f*
artifacts, 524–531
attenuation, 481–483, 481*t*, 482*f*
axial resolution, and image quality, 497–498
B-mode, transducer, 508–509
beam former, 502
beam steering, 505
biological mechanism, effect, 551–552, 553*f*
contrast sensitivity, noise, and image quality,
525–526
damping block, 486–487, 487*f*
Doppler performance measurements, 548
dynamic focusing, 505
electronic scanning and real-time display,
512–515, 512*f*, 513*f*
elevational resolution, and image quality, 500
enhancement artifact, 527*f*, 529
grating lobe artifact, 530
harmonic imaging, 517–522, 518*f*, 519*f*, 520*f*,
521*f*
image data acquisition, 501–510
image quality, 524–531
interaction with matter, 476–483
lateral resolution, and image quality, 498–500
M-mode, transducer, 509, 509*f*
matching layer, 487–488
measurements area, distance, volume, 516, 516*f*
mechanical scanning and real-time display,
510–511, 511*f*
mirror image artifact, 530
multipath reflection artifact, 530
phased array, 490
piezoelectric crystal, 483–484, 485*f*, 486*f*
preamplification, 504–505, 505*f*
pulse echo operation, 502–504
pulse echo technique, 13
pulser, 502
quality assurance, 544–548, 545*f*, 546*f*, 548*t*
receiver, 505–508, 506*f*, 507*f*
reflection, 477–479, 478*f*, 479*t*
refraction, 479–480
resonance transducers, 484–486
reverberation, 528*f*, 529–530
scan converter, 509–510

scattering, 480–481, 480*f*
shadowing artifact, 527*f*, 529
side lobe artifact, 530
signal summation, 505
slice thickness, 531
sound, characteristics of. *See* Sound,
characteristics of
spatial resolution, 524–525
special purpose transducer assemblies, 522, 523*f*
speed artifact, 528*f*, 530
system performance, 544–548
three-dimensional imaging, 522–524, 524*f*
transducer, 12, 483–490, 484*f*
transmit/receive switch, 502
two-dimensional image display and storage,
510–516
ultrasound contrast agents, 517
wave interference patterns, 474*f*

Ultrasound beam

far field, 493
Fraunhofer region, 493
Fresnel region, 490–493
grating lobes, 495–497, 496*f*
Huygen's principle, 490
nearfield, 490–493, 491*f*, 492*f*
properties of, 490–500, 491*f*
side lobes, 495–497, 496*f*
spatial resolution, 497–500

Ultrasound imaging, 12–13, 12*f*Uncharged particle interaction, 36–37, 36*f*Units, radiation. *See* Radiation quantities

U.S. Food and Drug Administration

radiopharmaceuticals, regulation, 624

U.S. Nuclear Regulatory Commission, 790

radiopharmaceuticals, regulation, 625

regulations, ALARA principle, 792

Useful beam, sources of exposure, radiation
protection, 759**V**

Valence shell, 22, 23

Variable frame rate pulsed fluoroscopy, 244–245

Varying magnetic field effects, biologic effects
magnetic resonance imaging, 466

Vector *versus* scalar quantities, 865, 866*f*

Veiling glare, output phosphor, 236

Velocity, 532, 867

Video camera, fluoroscopy, 240–242, 240*f*

Video interface, 77

Voltage, Bremsstrahlung spectrum, 97

Voltage ripple, x-ray, 131–132, 132*f*

Voxel volume,

computed tomography, 329

magnetic resonance imaging, 440

W

WAN. *See* Wide area network

Water spots, processor artifacts, 182

Wavelength
 electromagnetic wave, 17
 sound, 471-475, 473f
 Wet pressure marks, processor artifacts, 182
 Wide area network (WAN), 555
 Windowing
 contrast enhancement, 90-91, 90f, 91f
 defined, x-ray computed tomography, 358
 Word, digital data representation, 63
 Word mode, digital image formats in nuclear medicine, 696
 Wrap-around artifacts, magnetic resonance imaging, 455-457, 457f
 Write-once, optical disk, 75
 Written directive, radiopharmaceuticals, 625-626

X

X-ray

collimator, 114, 115f
 emission, factors affecting, 135-137, 136f, 137f
 exposure timing, 132-135
 electronic switch/timer, 132-133
 falling-load generator control circuit, 134, 135f
 mechanical contactor switch, 132-133
 photomultiplier tube, 133
 phototimer switch, 133-134, 134f
 filtration, 114, 115t
 generator, 116-124
 components, 116-123, 123f
 autotransformers, 120-121, 120f
 diodes, 121-122, 121f, 122f
 filament circuit, 126
 rectifier circuit, 124-126, 125f
 transformers, 116-119, 117f, 118f
 exposure timing, 132-135
 line voltage compensation, 119
 phototimer switch, 133-134, 134f
 types, 126-132
 constant potential generator, 129
 inverter generator, 129-131, 130f
 single phase x-ray generator, 126-127, 126f, 127f
 three-phase x-ray generator, 127-129, 128f
 voltage ripple, 131-132, 132f
 positive beam limitation, 116
 production, 97-102, 98f
 bremsstrahlung process, 97-100, 99f
 x-ray spectrum, 100-102, 100t, 101f, 102t
 tube, 102-112, 103f
 anode, 105-112, 106f
 angle, 108-111, 108f
 configuration, 106-108, 106f

focal spot size, 108-111, 108f, 111f
 heel effect, 112, 112f
 off-focus radiation, 112
 pinhole camera, 110
 rotating, 106, 107f
 slit camera, 110
 star pattern, 110
 stationary, 106
 biased focusing cup, 104, 104f
 cathode, 102-105, 103f
 cathode block, 104
 getter circuit, 113
 housing, 113
 grid biased tube, 113
 manimography tube, 113
 insert, 113
 leakage radiation, 113
 rating chart. *See* Rating chart, x-ray tube
 space charge effect, 104
 thermionic emission, 103
 X-ray computed tomography, 6, 7f, 327-372
 acquisition, 329, 330f
 artifacts, 369-372
 basic principles, 327-331, 328f, 329f
 beam hardening, 369-371, 370f, 371f
 detector type, 339-342, 339f
 digital image, 358-367
 fan beam geometry, 329
 filtered backprojection reconstruction, 330, 353-355, 354f
 helical scanner, 337-338, 350-351, 350f, 351f
 Hounsfield unit, 356-357
 iterative reconstruction, 351
 motion artifact, 371
 multiplanar reconstruction, 358-360
 parallel beam geometry, 329
 partial volume effect, 371-372, 372f
 radiation dose, 362-367
 simple backprojection, 351-353, 352f
 slice thickness, 342-344, 343f, 344f
 slip-ring technology, helical scanner, 337
 stack mode viewing, 362
 third generation, 333-334, 334f
 fourth generation, compared, 336, 336f
 three-dimensional image display, 360-362, 361f
 tomographic reconstruction, 330-331, 331f, 346-358, 347f, 348f
 window/level control, 358, 359f, 360f
 windowing, defined, 358
 X-ray intensifying screen, 150-153
 X-ray spectrum, 100-102, 100t, 101f, 102t
 Xenon detector, x-ray computed tomography, 339-340, 339f





"Nope ... no sign of your kitten, Ma'am.
But to be absolutely certain, we'd better
perform a CAT scan."

RUBES*

Creators Syndicate, Inc.
© 1993 Leigh Rubinf

The Essential Physics of Medical Imaging

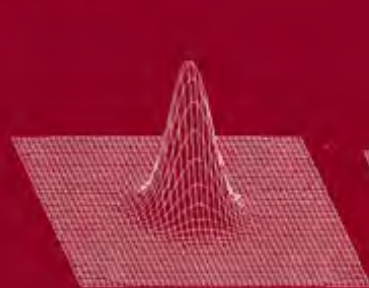
Second Edition

Jerrold T. Bushberg, Ph.D., J. Anthony Seibert, Ph.D.
Edwin M. Leidholdt Jr., Ph.D., John M. Boone, Ph.D.

The classic textbook gets even better, as scientists/educators Bushberg, Seibert, Leidholdt and Boone have completely updated this volume to include all aspects of modern medical imaging. This text is sure to become the new teaching standard for medical imaging professionals. Relying on their decades of experience teaching radiology residents, the authors help the reader to easily understand the theory and applications of the science of medical imaging, including:

- Basic Concepts
- X-ray Imaging
- Ultrasound
- Magnetic Resonance Imaging
- Nuclear Medicine
- Radiation Protection
- Radiation Dosimetry
- Radiation Biology

The Essential Physics of Medical Imaging offers more than 570 detailed illustrations, 125 tables, and five appendices that provide a review of the principles of physics related to medical imaging, and other useful reference data and resources. A significant revision and expansion, this edition is an invaluable learning guide and an essential reference on the bookshelves of radiologists, biomedical engineers, and medical physicists.



PSF



LSF



ESF

